CRM 97-014 / March 1997

Understanding Aircraft Readiness: An Empirical Approach

Laura J. Junor • James M. Jondrow •
Peter Francis • Jessica S. Oi

Center for Naval Analyses
4401 Ford Avenue • Alexandria, Virginia 22302-1498

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **MAR 1997**	2. REPORT TYPE	3. DATES COVERED **00-00-1997 to 00-00-1997**
4. TITLE AND SUBTITLE **Understanding Aircraft Readiness: An Empirical Approach**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **CNA,Center for Strategic Studies,4825 Mark Center Drive,Alexandria,VA,22311**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**94**	

Approved for distribution: March 1997

Samuel D. Kleinman, Director
Infrastructure and Readiness Team
Support Planning and Management Division

CNA's annotated briefings are either condensed presentations of the results of formal CNA studies that have been further documented elsewhere or stand-alone presentations of research reviewed and endorsed by CNA. These briefings represent the best opinion of CNA at the time of issue. They do not necessarily represent the opinion of the Department of the Navy.

Contents

Summary

In the wake of a changing defense climate, the Navy is continuing to find ways to adjust to its smaller size while maintaining its ability to respond when required. An important part of the strategy is to monitor readiness during the downsizing process.

The first step toward managing readiness is to understand what readiness is and why it changes over time or among units. This paper contributes to the further understanding of readiness by identifying the relationship between standard readiness measures and their determinants for Navy fighter (VF), attack (VA), and fighter/attack (VFA) aircraft.[1] The analysis is an extension of our earlier work on explaining the readiness of surface combatants.[2] Our objective was to build a comprehensive database of Navy fighter and attack units over time and identify readiness trends and relationships between readiness determinants and readiness measures where they exist.

We analyzed unit-level readiness data for VA, VF, and VFA squadrons.[3] We built a set of equations around three readiness resource areas: personnel, equipment, and unit training. Using regression analysis, we are able to demonstrate the impact of a change in the determinants.

We find that:

- The readiness of Navy fighter and attack units shows evidence of a peak in the early 1990s and a low point in the early 1980s.

1. The work is sponsored by the Deputy Undersecretary of Defense for Readiness and N81.
2. See Robinson et al., 1996, or Junor and Oi, 1996.
3. The data are for individual squadrons, monthly from January 1982 through December 1995.

- Current readiness is generally high, but suffers from unit training problems.
- Fully mission capable (FMC) rates and Status of Resource and Training (SORTS) equipment scores measure different dimensions of aircraft equipment condition and should not be used interchangeably.
- Different type/model/series (TMS) of aircraft have significantly different readiness profiles. For example, the F/A–18 typically has relatively low scores for training but performs well in the areas of personnel and equipment condition.
- Personnel quality remains an important driver of readiness.
- Training depends on FMC flight hours.
- The readiness of a squadron depends, in part, on the readiness of its carrier.[4]

The paper includes sections on:

- The output measures of readiness and why we chose them. The discussion includes a description of the data and of changes over time.
- Our analytical method and how our model is laid out.
- Our results, equation by equation—including both a theoretical discussion of our prior expectations and estimates from the data.[5]
- Our personnel quality index for aircraft squadrons.[6] Using this index, we can demonstrate the sizeable impact of the increase in squadron enlisted quality on readiness. The results indicate

4. We could not find a statistical relationship between the readiness of a squadron and its base.
5. Actual regression coefficients and technical notes are reserved for the appendices.
6. Readers familiar with the index as applied to surface combatants will find many similarities.

that enlisted quality is an important asset not only for personnel readiness but for FMC rates and ultimately for training readiness as well.

- A synopsis of our findings and how they could be used to help manage readiness.

THIS PAGE INTENTIONALLY LEFT BLANK

Measures of readiness

We measure readiness using historical data from SORTS and data on FMC rates.

SORTS

We used the Status of Resources and Training System (SORTS) scores to measure readiness in the areas of personnel and training.[7] Readers unfamiliar with SORTS can think of it as a grading system for the status of resources in four areas: personnel, supply, equipment, and training. It captures only some aspects of readiness (focusing on resource sufficiency and completion of training events). It is often used as a proxy for readiness more generally and does incorporate some of the commanding officer's opinion on the readiness of the unit.

SORTS scores are computed by all units in all services. Each unit computes a series of ratios for each resource area that measures the quantity of a resource compared to what the ship was designed or authorized to carry. These percentages are transformed into ordinal scores in each resource area.[8] The scores range from C1 to C5, with C1 being the highest score. Typically, units are at least C2 to deploy. C5 is reserved for squadrons undergoing depot maintenance leaving C4 as the lowest operating score. Once scores for each resource area

---

7. We do not explain squadron supply readiness in this paper. Because the squadron's supplies are so closely linked to (or controlled by) the hosting base or carrier, we regard supply as being exogenous to squadron readiness. In other words, the impact of measures of supply readiness are important to squadron readiness, but they are determined outside the squadron.

8. Scores are also computed by primary mission area (e.g., mobility and antiair).

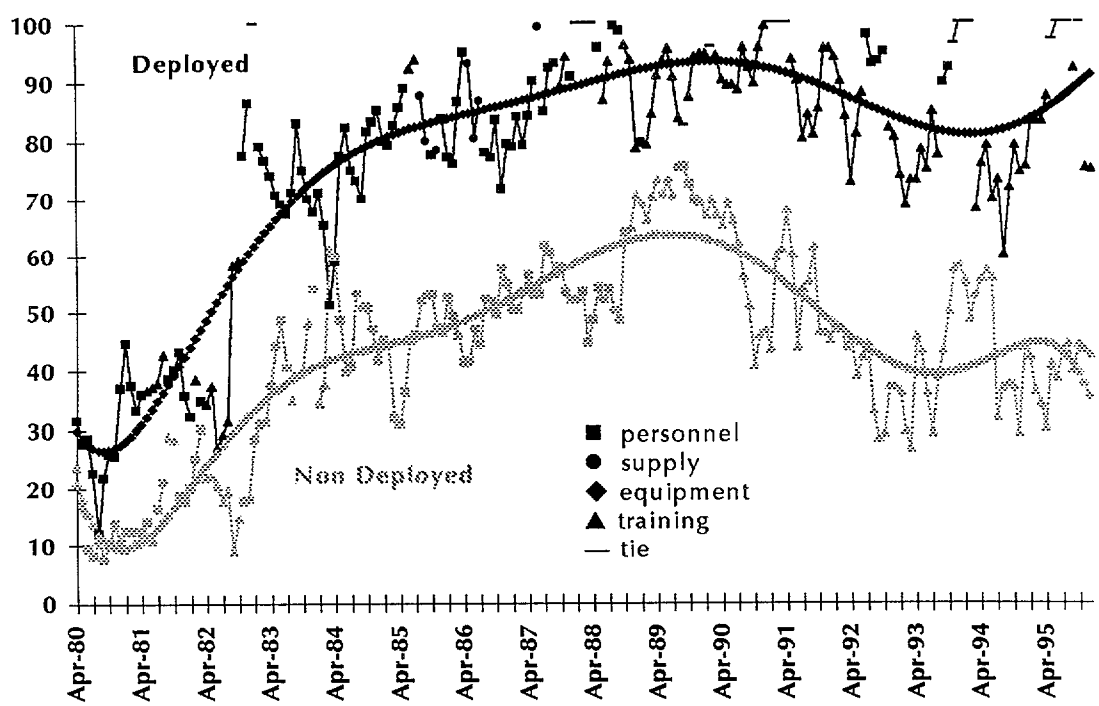
have been computed, the squadron takes the lowest of these scores and uses it as an overall measure of the squadron's resources. Unit-level scores are reported to the type commanders, fleet commanders, unified commanders, and, ultimately, the National Command Authority (NCA). Units must submit revised scores within 4 hours of a change in status.

We used SORTS scores as our readiness proxy for personnel and training because it was the only widely used measure of readiness and the only one available at the squadron level going as far back as the early 1980s.

SORTS trends since 1980

The first step toward understanding squadron readiness is to see how it changes over time. These changes are shown in figure 1.

Figure 1. Overall SORTS for fighter and attack squadrons



This chart shows a time series of overall SORTS scores for active duty VF, VA, and VFA squadrons. Each point represents the average percentage of time squadrons spent in either C1 or C2 in a month from 1980 through 1995. The data are for deployed and nondeployed squadrons. The continuous black lines through both the deployed and nondeployed series are a smoothed version of the data. The purpose is simply to help identify general trends.

The overall SORTS score reflects the lowest score given in four resource areas: personnel, supply, equipment, and training. The figure identifies the weakest resource areas over time. The height of the marker represents time in C1 or C2; the type of marker reflects the driving resource area. The problem areas are almost always personnel and training.

A notable characteristic of these data is the variability. Movements from month to month are extremely noisy and offer little hope for accurately predicting short-run movements in readiness. Even attempts to smooth the data sacrifice a lot of the information (as seen by the amount of data distant from the solid line). This means that, at best, we will probably be able to predict or explain broad, general movement in readiness.

Generally, the data show a growth in readiness over 15 years. The data also reflect at least a portion of the Navy's hollow period following the Vietnam War—a period usually considered to begin in the late 1970s and extend to the early 1980s. Our data only start in 1980, but we see that these units were at their 15-year low in early 1981. On average, deployed squadrons were mission ready only 30 percent of the time. Clearly, a deficiency in personnel readiness was a leading cause of the problem. Later, we will see that personnel quality may have been the dominant problem.

Over time, the Navy pulled out of this trough and appeared to be at peak readiness around 1989. There was a rather long period of relatively high readiness spanning about 8 years. At that point, training issues quietly replaced personnel as the weakest resource area. From 1989 until mid 1995, overall SORTS, still led by training, was falling. Although a few months of data suggest that the Navy has begun to recover on the deployed side, nondeployed scores were still declining at the end of 1995. The reason codes behind these particular data are often not informative. One of the most common reasons cited for not being C1 in training during the last 5 years is "other." More informative reasons suggest problems accomplishing or completing training either because there was no access to a training area or because of administrative or operational commitments. A lack of necessary equipment (including weapons) is also a common reason for the decline.

Deployed and nondeployed squadrons appear to move together throughout most of these series. Only in the last several months have we seen any evidence suggesting that the nondeployed squadrons are being sacrificed for the deployed squadrons. Because of the limited number of observations hinting at this possibility, it is too early to tell if this is happening.

Fully mission capable rates

We found that fully mission capable (FMC) rates were the best single measure of equipment condition. An aircraft FMC aircraft is one that is able to perform all its missions. FMC rates are computed for three different levels: for individual aircraft, as daily spot checks for the squadron and as a percentage of time for the squadron. Here we use the percentage of time that the aircraft in a squadron are able to perform all of their mission areas.

There are many common measures of aircraft equipment condition: equipment SORTS, fully mission capable (FMC), mission capable (MC), partially mission capable due to supply (PMCS), partially mission capable due to maintenance (PMCM), not mission capable (NMC), not mission capable due to supply (NMCS), and not mission capable due to maintenance (NMCM).

To determine the best single measure of equipment condition, we applied an indexing technique, called principal component analysis, to these variables.[9] The first of the two indexes explains 80 percent of

9. Principal components is a mathematical means of forming indexes from a group of variables. It has the property of accounting for the maximum possible amount of variation in the data, using a simple weighted average of the variables.

the information in the data. Upon plotting the data, we realized that this weighted average was nearly 99 percent correlated with the FMC rate. This suggests that of the seven original variables, FMC is the most representative of the information. For details, see appendix A.

The second principal component, or index, appears to be almost entirely weighted on equipment SORTS. This suggests that FMC rates and equipment SORTS systematically measure different dimensions or aspects of squadron readiness. Upon further investigation, we learned why this might be true:

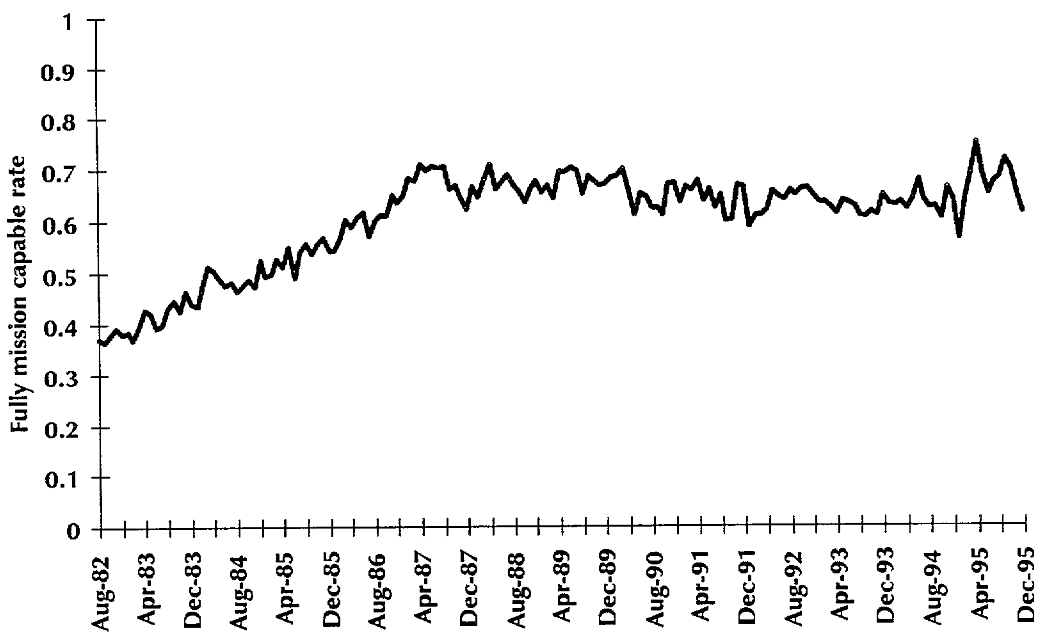
- *Equipment SORTS measures the number of available aircraft relative to M+1 requirements.*[10] Its purpose is to monitor whether the squadron has a sufficient number of capable aircraft. Because of the frequency of scheduled and unscheduled maintenance actions, *SORTS does not require maintenance actions to count against the readiness status as long as the aircraft is expected to be operational within 24 hours.*
- *FMC rates measure the percentage of time that aircraft are fully able to meet the missions they are required to meet.* This measure is meant to be an accurate portrayal of the availability of mission ready aircraft. *Therefore, it includes downtime associated with all maintenance actions.*

FMC trends since 1980

FMC rates indicate that equipment condition for tactical aircraft is fairly strong throughout our data period (see figure 2). Data from the early 1980s show strong growth until late 1986. At first, we thought that this growth reflected the introduction of the new, more reliable F/A–18s; however, we saw similar growth patterns for the F–14s. Since 1986, the FMC rates varied around 65 percent but did not show a pronounced upward or downward trend.

10. M+1 requirements refer to the resources necessary the first day after mobilization for a major contingency.

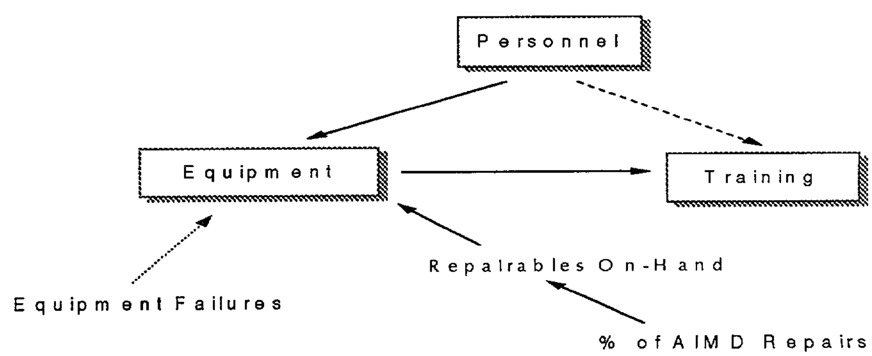
Figure 2. Historical FMC rates for active duty VA, VF, and VFA squadrons



Modeling

In this paper, we are trying to understand the relationship between determinants of readiness and standard measures of readiness in the resource areas of personnel, equipment, and training. In some cases, measures of readiness in one resource area are determinants of readiness in another area. The following diagram illustrates some of the relationships that we explored in our analysis.[11] Personnel readiness, for example, could influence equipment readiness (FMC rates) and unit training. The impact on unit training could be direct, or indirect, through equipment readiness.[12]

Figure 3. A theoretical model of readiness



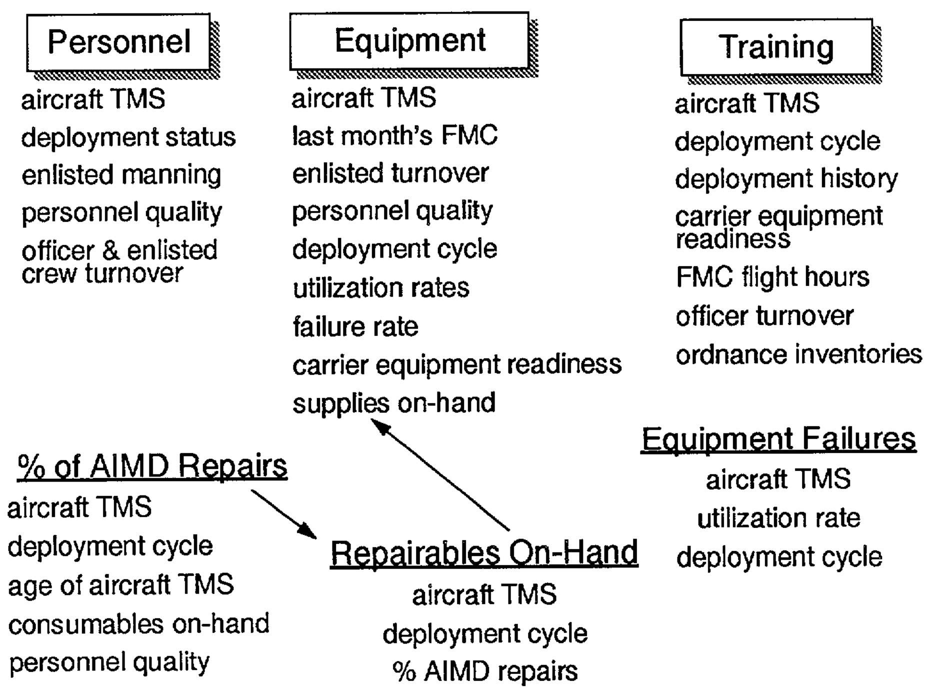
We test the hypotheses involving determinants and measures of readiness by building a group of equations around the measures. We

11. Dashed lines indicate relationships that we believe are theoretically possible but were unable to substantiate.

12. Cavalluzzo and Horowitz (1987) present the interrelationships between readiness and resource areas.

use regression analysis to identify relationships. Figure 4 represents this group of related equations. The equations contain three supporting equations for equipment condition: an equation explaining the number of failures relative to flight hours, and two others explaining the relationship between onsite repairs, the inventory of repairables, and equipment readiness.

Figure 4. An illustration of the readiness equations



One can imagine a similar system of equations with arrows pointing in all directions, so that each equation feeds into the next. This type of system would present great difficulties in sorting out relationships. In the present case, our knowledge of the process and initial estimates suggested that the structure was much simpler.[13] For example, we felt that equipment condition was an input to unit training, but that unit training was not an input to equipment condition.

---

13. Cavalluzzo and Horowitz's theoretical relationship among resource areas tend to support this claim.

Readiness equations

In this section, we first present our prior expectations about the effects of particular readiness determinants. We then turn to measures and empirical evidence of effects. Subsections deal with individual resource areas, starting with personnel.

Personnel

As our measure of personnel readiness, we use the fraction of time the squadron spent in C1 for personnel according to SORTS.[14] This measure is designed to identify the extent to which the squadron has enough personnel and the ability to allocate those personnel to specific jobs (where the allocation is based on factors such as seniority, job speciality, and training).

With this in mind, we would expect variables reflecting the quantity of enlisted and officer personnel to have a positive impact on personnel readiness.[15] We would also expect variables reflecting personnel turnover to have an adverse effect. There are two reasons for this:

- High turnover suggests that many people do not have a lot of experience with that particular unit. One could imagine a mechanic with only A–6 experience reassigned to an *F/A–18* squadron. He may be quite capable, but there will be a learning curve associated with his arrival.

14. We have chosen to measure SORTS as percentage of time in C1 rather than C1 or C2 because C1 is probably less gamed than C1/C2. The rational is that commanding officers are more likely to game from C3 to C2 (the lowest you can still deploy) than from C2 to C1.

15. The variable measuring the quantity of enlisted personnel is actually a ratio of available personnel (weighted by pay) to required personnel. Weighting personnel by pay ensures that losses of senior personnel are felt more strongly than losses of junior personnel.

- High turnover takes away from unit cohesion. Although SORTS does not measure cohesion directly, it may sway a commander's assessment of his unit's readiness.

The quality of enlisted personnel should also have a positive influence on personnel readiness despite the fact that it isn't directly measured by SORTS reporting. Brighter, more experienced personnel may be more likely to have skills that match the unit's requirements and may be more easily cross-trained. Furthermore, commanders may subconsciously consider quality to be a substitute for quantity at the margin.

Other factors may help explain differences between personnel readiness levels either between squadrons or over time. The first is the type of aircraft within the squadron. Some aircraft types, most notably the F/A–18, simply have different manning requirements than their older counterparts. We also expect readiness to change over the course of a deployment with deployed squadrons being the most ready and nondeployed showing an improvement in readiness as their next deployment approaches.

Estimation method

We tested for the above relationships by using regression analysis to estimate equation 1 below.[16]

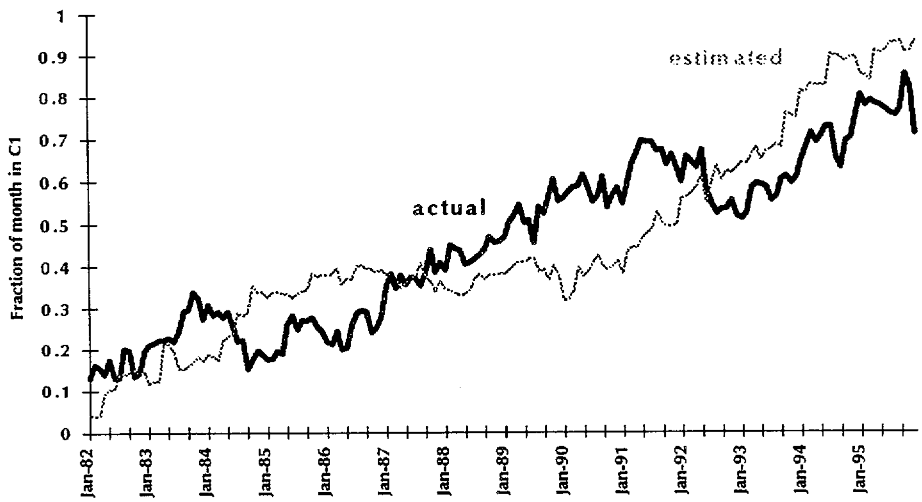
Personnel = P(enlisted manning, enlisted turnover, enlisted personnel quality, officer turnover, deployment indicator, TMS indicators) (1)

The equation seeks to explain the fraction of time that squadrons are C1 for personnel each month. Figure 5 illustrates how the average actual value matches with the average predicted value over time. While our estimates are close toward the end, we only pick out the general trend for most of the time period.[17]

16. We used a logistic equation to estimate personnel readiness.

17. Appendix B contains the original coefficients and a technical explanation.

Figure 5. Estimated versus actual personnel readiness



Estimation results—the impact of personnel drivers

As expected, we find that measures of the quantity and quality of crew personnel significantly affect the fraction of time a squadron is C1 for personnel. Table 1 summarizes the effect of the determinants. The table shows how the percentage of time that a squadron spends in C1 changes in response to a specified change in one of the determinants.

Among the determinants, the personnel quality index (for enlisted personnel) had the strongest impact on personnel readiness. Because this variable is measured as an index, there is no intuitive interpretation of what a 10-percent change means. We can, however, take advantage of the historical aspect of our data and use benchmarks (historical highs and lows) to illustrate the impact of personnel quality on readiness. The next section expresses the impact of personnel quality in this way. Not surprisingly, squadron manning is also positively associated with personnel readiness. We see that a 10-percent increase in the number of enlisted personnel relative to requirements causes a 2-percent increase in time in C1. Other statistically significant drivers are officer and enlisted turnover; however, their impact is very small. A 10-percent increase in either only causes time in C1 for personnel to drop by a few hours per month.

Table 1. The impact of drivers on personnel readiness

Significant drivers	Resulting change in personnel readiness
Deployed relative to nondeployed	7%
A–7E relative to A6–E	65%
F–14A relative to A6–E	5%
F–14B relative to A6–E	44%
F–14D relative to A6–E	23%
F–4S relative to A6–E	1%
F/A–18A relative to A6–E	84%
F/A–18C relative to A6–E	62%
An increase in enlisted personnel quality	a
A 10% increase in manning	2%
A 10% increase in officer 3-month turnover	-0.2%
A 10% increase in enlisted 3-month turnover	-0.6%

a. The impact of personnel quality is reserved for the following section.

Aside from the impact of personnel quality, the deployment status and the type of aircraft seem to be the most important explanatory variables in this equation. A switch from nondeployed to deployed status increases the amount of time that squadrons are ready by about 7 percent or 2 days per month. We also see that F/A–18 squadrons seem to be significantly more likely to be ready than the other types of aircraft.[18]

Equipment condition: flight hours between failures

Flight hours between failures reflect how often squadrons experience equipment failures. Generally, squadrons that have fewer failures (more flight hours between failures) would be more likely to be equipment ready.[19]

18. We used the A–6s as the omitted class in each regression. The coefficients in the appendix are then interpreted relative to the A–6s.

19. Unfortunately, when we estimated the FMC equation, we did not see this. We will discuss this finding more in the FMC section.

A fundamental determinant of flight hours between failures is the type of equipment. For example, the F/A–18 was designed to be particularly reliable. Another basic determinant should be measures of use such as utilization rates or the number of sorties per aircraft. Furthermore, we would expect different effects depending on whether the planes were in an embarked environment.[20]

The place in the deployment cycle may also help explain differences in the quantity of failures over time or between squadrons. Holding the effect of carrier landings constant, deployed squadrons should be more ready, experience more flight hours between failures, than non-deployed squadrons. Nondeployed squadrons should show an improvement in the number of failures as their next deployment draws near. The rationale is that squadrons take advantage of not being deployed to correct problems incurred during the last deployment.

Estimation method

We estimate the number of flight hours between failures monthly for each squadron using equation 2 below.

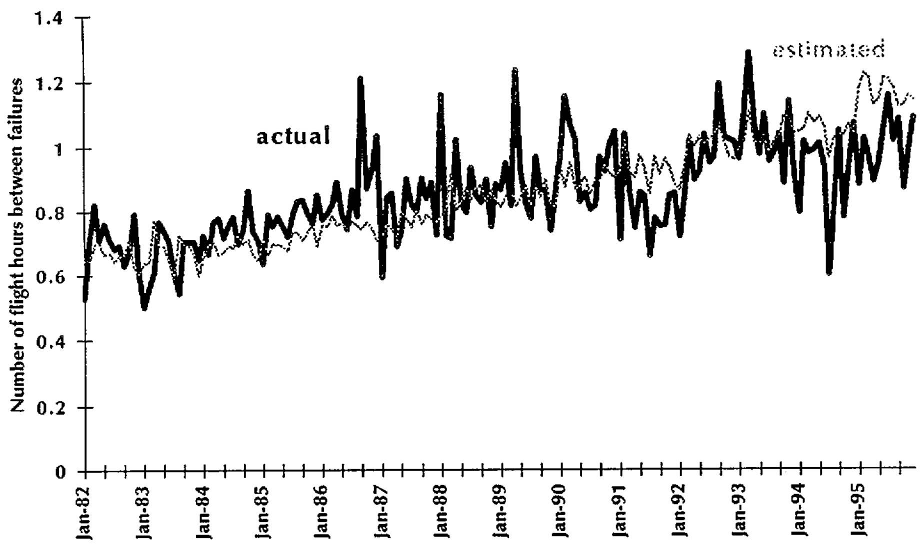
$$\textit{Flight Hours Between Failures} = F(\textit{utilization, months since last deployed, sorties, TMS indicators}) \qquad (2)$$

The determinants, taken together, provide significant explanatory power. Figure 6 shows how well average predicted values match average actual values over time. Again, we appear to explain the general trend in failures relative to flight hours, but we do not replicate the high variance of the actual data.[21]

20. An embarked environment is notoriously hard on equipment because of landings, exposure to salt water, confined areas, and a complete dependence on the carrier.

21. We tried estimating separate equations for each aircraft type in order to reduce the noise and increase our explanatory power. Generally, these equations did not yield significant coefficients.

Figure 6. Actual versus estimated flight hours between failures



Estimation results—the effect of drivers on flight hours between failures

The number of failures between flight hours appears to be almost entirely a function of the type of aircraft, its position in the deployment cycle, and how often the aircraft are being used. Table 2 summarizes the impact of the statistically significant variables.

The actual impact of moving from nondeployed to deployed status does not have an important impact on failures despite the fact that it is statistically significant. As the months pass since a squadron was last deployed, however, there are significantly more flight hours between failures. We believe this reflects a few things. First, just after a deployment, aircraft commonly undergo a lot of intensive maintenance. Often this level of maintenance is accompanied by an increase in the frequency of failures as problems are being worked out. Second, approaching deployment causes an increase in the amount of flight hours as the unit trains. If failures increase less proportionately with flight hours, we could merely be seeing the number of failures diluted by a significant rise in flight hours. Finally, the impact of the

Table 2. The impact of drivers on flight hours between failures

Significant drivers	Resulting changes in the number of flight hours between failures
Deployed relative to nondeployed	0.002
A–7E relative to A6–E	0.19
F–14A relative to A6–E	0.03
F–14B relative to A6–E	-0.05
F–14D relative to A6–E	0.36
F–4N relative to A6–E	0.23
F–4S relative to A6–E	0.20
F/A–18A relative to A6–E	0.83
F/A–18C relative to A6–E	-1.10
A 10% increase in the months since deployment	0.006
A 10% increase in the number of sorties per aircraft	0.01
A 10% increase in the utilization rate	0.02

deployment status may be captured by the sortie and utilization variables—deployed planes fly more than nondeployed planes.

We did find a significant relationship between the number of sorties flown per aircraft and the flight hours between failures. We found that, overall, an increase in the number of sorties flown has a positive effect on flight hours between failures. However, there does appear to be a difference when we divide sorties by deployment status. In fact, we see that deployed sorties have a significant negative impact on flight hours between failures.[22] This probably reflects the trauma that the carrier environment inflicts on the aircraft. The utilization rate of aircraft (the number of flight hours flown per aircraft) also has a positive impact on flying time between failures. A 10-percent increase yields a slight increase of 0.02 in flight hours between failures. This result is independent of deployment status.

22. The number in the table represents the total effect sorties has on flight hours between failures. Note that the sign is positive, implying that the deployed effect is smaller than the nondeployed.

The final set of variables reflects the type of aircraft in the squadron. The important thing to note here is that there are significant difference between the types and generations of aircraft. For example, there are interesting differences between the two series of F/A–18s. Switching to the F/A–18A from some aircraft other than an F/A–18 increases the number of flight hours before failure by 0.83. The same switch for the F/A–18Cs, however, decreases flight hours between failures by more than one.

Equipment condition: the proportion of AIMD repairs

The proportion of repairs logged at the Aviation Intermediate Maintenance Department (AIMD) that are actually done at the AIMD level is important to readiness because of the sizeable cost inherent in sending the "black box" to the depot.[23] On the surface, it seems plausible that readiness should be improved by an AIMD that is clever enough to do its own repairs, thereby bypassing the slow depot process. Given that a part requires a repair, it would seem that the more often the repair was done locally, the better the stock of working repairables, and the higher the FMC rate for that squadron.

Those things that should theoretically affect the AIMD would include the quality and quantity of the ship and squadron personnel. The theory is that an adequate supply of talented mechanics may find ways to fix parts locally or perform better preventative maintenance, leaving all but the most extreme cases for the depot level.

The equipment readiness of the carrier or base should ultimately have an impact on its ability to repair items. This equipment ranges from the highly technical test benches to the lower-tech cranes and elevators. To repair items locally, these pieces of equipment must work properly. We also expect that the availability of supplies is important in allowing onsite repairs.

The other primary set of variables relevant to the proportion of AIMD repairs is the TMS of the aircraft. There are major differences in the

23. Note that this variable refers only to repairs sent to the AIMD and thus will not include organizational-level activity.

way some of the aircraft represented in this analysis are constructed. Specifically, the F/A–18s are made of modular components, most of which are computer driven. When a part goes bad, the aircraft is designed so that the entire broken "black box" can be removed and replaced by a working one. This simple "remove-and-replace" procedure often takes place at the organizational level. Often the damaged black box is then sent to the depot. The total downtime is minimal, but probably a higher percentage of repairs are sent off site. In contrast, the F–14 uses older technology, and repairs can be done by the local mechanics.

We also expect to see an effect when a new series is introduced. A new series often has the support of the contractor when it first comes into the fleet. Because these contractors take care of most all repairs, few items are sent to the depots. Once the contractors leave, the AIMD and squadron personnel must learn the new system, and we would expect to see evidence of this learning curve in the ratio of onsite repairs. As a plane ages, personnel become more familiar with its repairs, and, if other factors are held equal, the proportion of AIMD repairs rises.

Estimation method

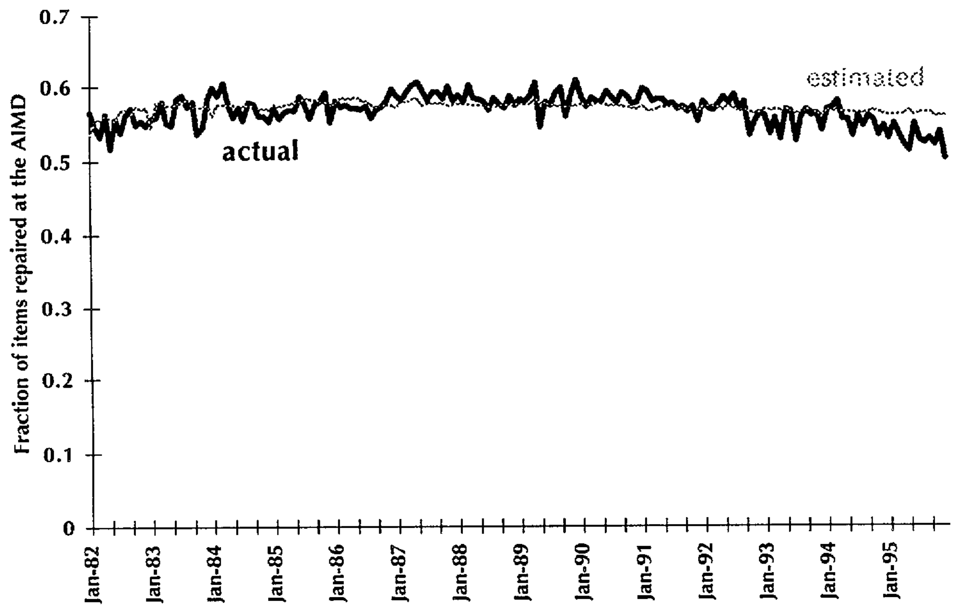
This equation explains the fraction of items sent to the AIMD that are actually repaired there.[24]

AIMD repairs = A(carrier personnel quality, squadron personnel readiness, deployment indicator, consumables on hand, the number of failures, age of the aircraft series, TMS indicators) (3)

Figure 7 indicates that the equation provides a fairly good fit for panel data. Once again, we have captured the general trend quite well, but miss on the variation in the original data. The determinants, taken together, have a significant impact on the repair rate.

24. See appendix B for technical details.

Figure 7. Actual versus estimated fraction of AIMD repairs



Estimation results—the effect of drivers on the proportion of AIMD repairs

We found that, in general, personnel readiness, the supply of consumables, and the TMS of the aircraft are the primary drivers of the proportion of onsite repairs per month. Table 3 lists our significant findings.

What we did not find was often as interesting as what we found. For example, we found no evidence that the readiness of the carrier or base influenced how many repairs were done locally. This may be because we included measures of the quality of those personnel and the amount of supplies on hand, which left little room for the carrier or base readiness to have an impact.

We also found that the personnel readiness of the squadron (measured as the number of days the squadron was in C1 for personnel)

Table 3. The impact of drivers on the proportion of onsite repairs

Significant drivers	Resulting changes in the proportion of items repaired at the AIMD
Deployed relative to nondeployed	-2%
A–7E relative to the A–6E	-0.2%
F–14A relative to the A–6E	2%
F–14B relative to the A–6E	3%
F–14D relative to the A–6E	-3%
F–4N relative to the A–6E	-5%
F/A–18 relative to the A–6EA	-3%
F/A–18C relative to the A–6E	-2%
A 10% increase in the amount of consumables on-hand	0.21%
A 10% increase in the number of failures	0.08%
A 10% increase in the age of the aircraft series	0.38%
An increase in the quality of the ship's enlisted crew	[a]

a. The impact of personnel quality is reserved for the following section.

had no significant effect on AIMD repairs. The personnel quality of the ship, however, had a very significant effect.[25]

The impact of the local supply of consumables was also surprisingly small. A 10-percent increase in this measure increased the proportion of onsite repairs by less than one percent. The number of failures had even less impact.

We found evidence of a learning curve with respect to new TMS. The longer aircraft have been around, the more likely their repairs will be done on site. Our hypothesis is that, over time, the AIMD personnel become more familiar with the aircraft and its failures and learn how to repair them.

25. Because a 10-percent increase in this index is not intuitive, we will evaluate the impact of personnel quality using a change in the index defined by historical benchmarks. We do this in the next section.

A major determinant of AIMD repairs is the TMS of the aircraft. The modular design and the computer-driven avionics of the F/A–18 appear to be the reason why we find that this type of aircraft is associated with fewer onsite repairs. In contrast, we see that the more conventional design of the F–14 is associated with more onsite repairs (with the exception of the F–14D which appears to be less susceptible to AIMD repairs).

Equipment condition: the percentage of repairables available within one day

We will demonstrate later in this section that the stock of repairables is an important determinant of the FMC rate. Theoretically, we also expect that more repairs done on site will increase the carriers's stock of repairables, thereby improving the FMC rate. This equation attempts to explain the availability of repairables and substantiates the link between onsite repairs and the FMC rate.

We will use the likelihood of fulfilling a request for a repairable within one day as a proxy for the stock of repairables. We have already discussed what is probably one important determinant of this probability—the proportion of AIMD repairs. We might expect that the stock of repairables would be responsive to the number of failures. Theoretically, the relationships could be either direct or indirect. On the one hand, a large number of failures could drain the ship's inventory. On the other hand, if a squadron (or TMS) has a notoriously high failure rate, we may see a compensating adjustment in the AIMD (stock of repairables) to maintain FMC rates.

Other influencing factors may be differences between TMS and the squadron's deployment status. Deployed squadrons or squadrons approaching deployment should have higher priority for replenishing inventories.

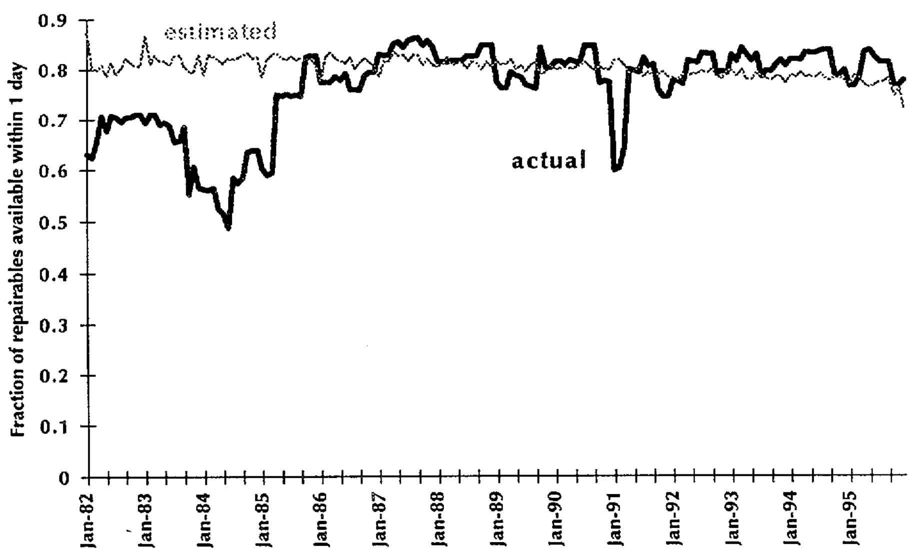
Estimation method

This equation explains the proportion of requests for repairables that were filled in one day or less.

Repairables on hand = R(proportion of items repaired at the AIMD, deployment indicator, months since last deployed, the number of failures, TMS indicators) (4)

Figure 8 compares our estimation with actual time series data. We are clearly overpredicting the early part of the data. This early decline in

Figure 8. Actual versus estimated repairables on hand



the number of on hand repairables may reflect a change in how repairables were financed. In the mid 1980s, the fiscal responsibility for repairables sent off site was shifted to the ship. This tended to make the ship more conscious about the relative costs of repairing items on rather than off site. One would expect then, that this policy change would increase the inventory of repairables, but that this change would work through the amount of items repaired locally. A second effect of this policy change was that it made the supply budgets more self-sustaining for the ships—and hence less susceptible to budget cuts. If placing the supply budget in the control of the carriers meant that they were actually better funded as a result, this may explain the increase in the availability of repairables. Future work on this equation should further explore this hypothesis.

The later part of the series indicates that we have captured the general trend, but once again leave much of the variation unexplained.

Estimation results—the effect of drivers on the proportion of repairables on hand

Generally, our results support our prior assumptions about influences on the stock of repairables. (See table 4.) One of the most important drivers is the proportion of AIMD repairs. A 10-percent increase in the number of repairs that are repaired locally increases the likelihood of having a repairable on hand when requested by 2 percent. We also see that deployed ships and ships approaching their next deployment appear to have more repairables on hand. Again, this supports the notion that deployed ships are able to maintain better inventories than nondeployed ships. We did find a weak, positive relationship between the average number of failures and the stock of repairables. Finally, there appears to be a difference between the availability of repairables between TMS. Relative to repairables for A–6s, those for F–14s and F/A–18s are less likely to be available within a day.

Table 4. The impact of drivers on the availability of repairables

Significant drivers	Resulting changes in the proportion of repairables available within one day
Deployed relative to nondeployed	11%
A–7E relative to the A–6E	1%
F–14A relative to the A–6E	-12%
F–14B relative to the A–6E	-0.1
F–14D relative to the A–6E	-11%
F–4N relative to the A–6E	-65%
F/A–18 relative to the A–6EA	-11%
F/A–18C relative to the A–6E	-12%
A 10% increase in the proportion of items repaired at the AIMD	2%
A 10% increase in the number of months since the squadron was last deployed	0.2%

Equipment condition: the FMC rate

Theoretically, we expect something to affect a squadron's FMC rate either because it reduces the likelihood of an equipment failure or because it increases the rate at which any repair is made. In this light, we expect four general categories of FMC determinants:

- Those regarding the quality or quantity of personnel
- Those contributing to the wear of equipment
- Those relating to the availability of repair parts
- Those reflecting the design of the aircraft.

We expect that the quality and quantity of personnel should not only improve the speed at which equipment is repaired, but may also reduce the likelihood of failures. You could easily think of a scenario in which having enough good people enables the squadron to perform excellent preventative maintenance.

Usage rates should also play an important role in the condition of a squadron's aircraft. Because planes are made to fly, we wouldn't expect the simple act of flying to adversely affect their condition. It would be excessive flying or flying under adverse conditions that would cause the problems. Another hypothesis revolves around the impact of the carrier environment on the aircraft. For example, carrier landings are often referred to as controlled crashes because of the force and strain that are placed on the aircraft (and the tailhook and landing gear). Thus, we should see deployed sorties having a negative relationship with FMC rates.

The availability of repair parts also influences the FMC rate. The simple theory is that the more supplies on hand, the faster repairs and maintenance can be made. Faster repairs and more maintenance easily translate into a higher FMC rate for the squadron.

The final category of influence reflects the design of the aircraft. Some TMS, namely the F/A-18, are designed with high FMC rates in mind. They are designed to fail less often. In addition, when a component breaks, the aircraft is made so that the "repair" is nothing more than a remove-and-replace of a black box. Assuming an adequate supply of black boxes (a repairable), the total downtime for this type of activity should be minimal.

Modeling the data efficiently

Although we often look at readiness as a snapshot (what fraction of our forces are ready right now), the data actually come in the form of durations (for example, how long is a particular unit C1 for personnel). This brings up the question of whether there is a way to explain the data that is truer to the form of the data and more efficient.

One way to do this is to use the duration in a particular readiness status as the dependent variable. For example, if a squadron was in C1 for training for 3 years, the value of the variable to be explained would be 3 years or 36 months. A difficulty in estimation is that the explanatory variables will be changing over this 3 years, so it is not clear what to use on the right-hand side of the equation. A number of indirect estimation techniques are available, but they are not very intuitive and are difficult to understand.

One technique does take advantage of the structure of the data in an intuitive and rigorous way. This technique is based on the similarity of duration data to the data modeled in a Markov process—so we will refer to the technique as the Markov technique.[26] In this project, we have illustrated the Markov technique on data for FMC, the fraction of time that a squadron is fully mission capable. By expressing current FMC as a function of the previous period's FMC, this technique identifies two important probabilities:

- The probability of moving from FMC to not FMC
- The probability of moving back.

26. A simple Markov process in one where the probability of any one outcome in a series of events is dependent on the outcome of the most recent event.

These probabilities are then related to explanatory variables. The technique has a number of attractive features:

- It can identify whether some variables have an effect on one probability but not the other. For example, it may turn out that quality of enlisted personnel is important in returning an aircraft to FMC status, but that officer quality is the key to remaining in FMC status.
- It can be applied to aggregated data. For instance, FMC applies to a single aircraft, but we only have data at the squadron level. The technique can be applied to the squadron-level data, whereas other techniques for handling durations cannot.
- It can be estimated with readily available statistical software for performing nonlinear regressions.

As a point of comparison for this technique, we also present the ordinary regression estimates. (See estimation method (1) below.)These specify FMC as a function of a number of explanatory variables and the value of FMC in the previous period.

Estimation method (1): FMC as a continuous process

We first report the more common approach where we simply model the percentage of the month that a squadron is FMC. This fraction is estimated for each squadron and for all VA, VF, and VFA aircraft. We used regression analysis to estimate the equation described below.

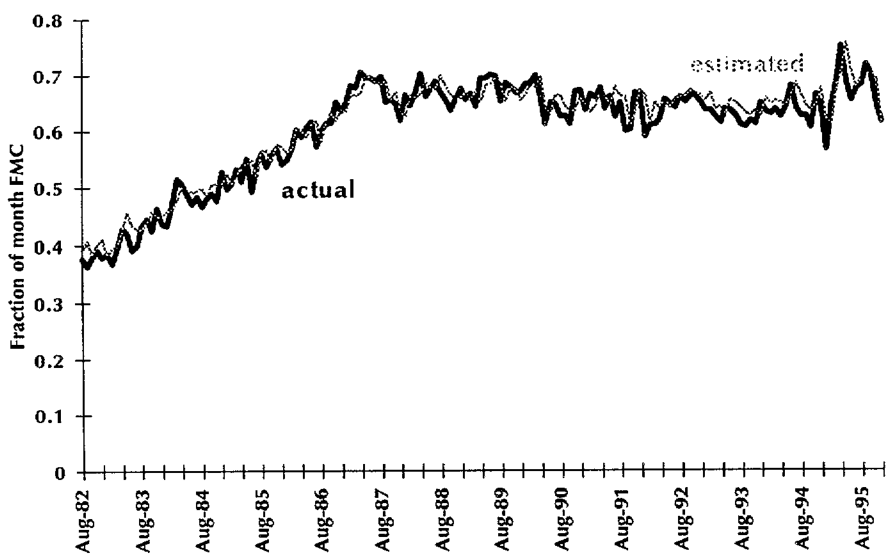
FMC = F(enlisted turnover, enlisted personnel quality, sorties, deployed sorties, supplies on hand, months since last deployed, current deployment status, the equipment readiness of the carrier, last month's FMC rate, TMS indicators) (5)

The model is estimated using a squadron's last month's FMC value as a determinant of this month's FMC. The reasoning here is that FMC movements appear to be adjustments from last month's value.[27]

27. This phenomenon is called autocorrelation in the econometric or statistical literature. One cause of autocorrelation is a sticky or sluggish process. If we believe that the movements in the variable are an adjustment from the last period, we can address the autocorrelation issue by including a lagged value of the dependent variable as an explanatory variable.

Figure 9 shows how the average actual values compare with the average predicted values over time. The actual and predicted values are very close over the entire span of the data.

Figure 9. Actual versus estimated FMC rate



Estimation results—the impact of drivers on the percentage of time squadrons spend FMC

As expected, we find that personnel, supply, and equipment indicators are significant drivers of the FMC rate. Table 5 lists the impact of the most significant variables.

Supply-related variables, although statistically significant, have only a modest impact on the squadron FMC rate. The equipment readiness of the carrier also plays a small role. When the carrier spends 10 percent more time in C1 for equipment, the squadron spends only a few additional hours in FMC per month.

We also found that personnel variables were statistically significant. Specifically, the quality of the enlisted squadron personnel is strongly

Table 5. The impact of drivers on FMC rates

Significant drivers	A resulting change in the FMC rate
Deployed relative to nondeployed	9%
A7–E relative to A6–E	-5%
F–14A relative to A6–E	-1%
F–14B relative to A6–E	-1%
F–14D relative to A6–E	-7%
F/A–18A relative to A6–E	4%
F/A–18C relative to A6–E	4%
A 10% increase in the amount of consumables on hand	0.3%
A 10% increase in the amount of repairables on hand	1%
A 10% increase in the number of days the carrier is C1 for equipment	0.1%
A 10% increase in the number of sorties per aircraft	0.5%
A 10% increase in enlisted 3 month turnover	-1%
An increase in enlisted personnel quality	a
A 10% increase in the months since the last deployment	1%
A 10% increase in the number of aircraft	-1

a. The impact of personnel quality is reserved for the following section.

linked to FMC rates. We will discuss just how much it affects readiness in the next section. Enlisted squadron turnover, as expected, has a negative impact on FMC rates. The size of the effect is modest with a 10-percent increase in turnover reducing the percentage of time in FMC by about one percentage point.

There is evidence of a deployment cycle. The process of moving from deployed to nondeployed status is associated with an increase in FMC of nearly 9 percent. That's an additional 3 days that the squadron is fully FMC per quarter. Furthermore, as the squadron moves farther from its last deployment, we find that it spends more days per month in FMC.

Related to the deployment effect is the effect of the number of sorties flown per aircraft per month. Generally, the more sorties flown, the higher the FMC rate. This result supports our assumption that machinery needs to be used to keep it running well. On the other hand, our results tell us that deployed sorties are associated with lower FMC rates. Again, this may indicate that carrier landings are inherently rough on equipment.

The TMS of aircraft also help explain a squadron's FMC rate. F/A–18s have significantly higher FMC rates relative to the A–6 (our base case), whereas F–14Ds have significantly lower rates.

Estimation method (2): FMC as a Markov process

The foregoing model of FMC rates is based on fairly standard empirical methods that are often used in applications where either:

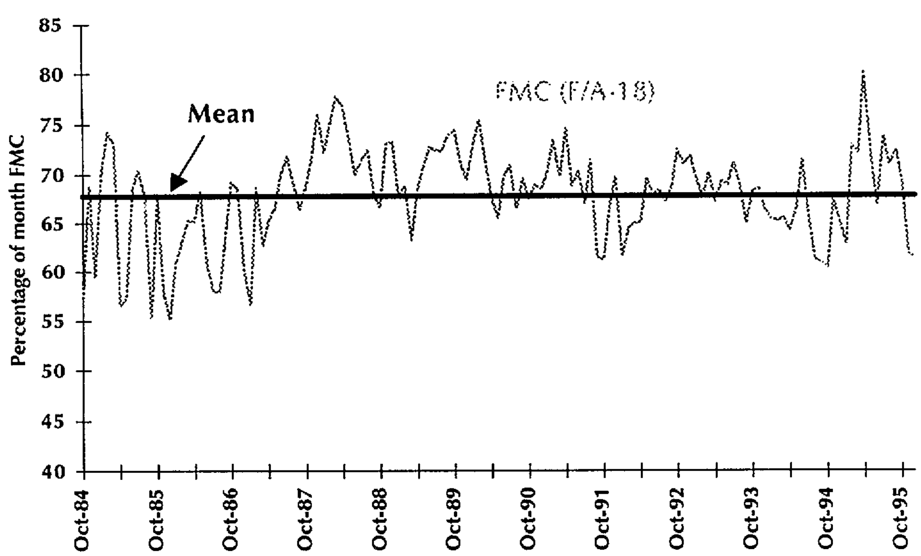
- There are no dynamic considerations
- Dynamics are present, but their exact form isn't known a priori.

As noted earlier, we have devoted some resources to exploratory development of a more structured model of FMC—i.e., a model in which the underlying process take on an assumed form. This alternative analysis is based on the concept of a Markov chain, which is a particular kind of stochastic process. The model that arises from this approach is more complex than what we get from the conventional analysis, and is therefore more difficult to work with. However, it has a couple of virtues when compared to the standard approach. These are:

- *Realism.* Model building, by its nature, entails making tradeoffs between realism and utility. All other things being equal, a more realistic model is to be preferred, both because it is more credible and because it is more portable—that is, it is more readily applied to new situations. The Markov chain approach makes use of some of what we know about the underlying mechanics of the process by which the observed FMC rates are generated. (Specifically, it accounts for the fact that the reported FMC rate represents an aggregation of information about individual aircraft.) Though still a simplification, it is arguably more realistic than a conventional linear model. As such, we would argue that using it is preferable to using a conventional model so long as—and this is a crucial qualification—the price that is paid in tractability is not too high. (More on the tractability question below.)

- *Accounting for observed autocorrelation.* Our initial motive for looking at a Markov-chain type of model was that we were searching for an explanation for a specific characteristic that FMC data routinely exhibit. That characteristic is *autocorrelation* (also called serial correlation). A series of data is autocorrelated if observations at different (typically adjacent) times are related. Figure 9 shows monthly FMC data for F/A–18s. (The horizontal line is the average over the relevant time span.) There are formal statistical tests to detect autocorrelation, and for the data in figure 9, they do indeed confirm that it is present. Less formally, the fact that there are several spells during which the FMC rates persistently remain above or below the mean is a signal that autocorrelation may be present. Such spells can be observed in figure 10.

Figure 10. F/A–18 data showing autocorrelation



Autocorrelation matters because the usual hypothesis tests are invalidated if the data are autocorrelated. There are ways of dealing with this problem in the usual linear modeling framework, but it is preferable to inquire whether a model of the underlying dynamics might give us insights into the cause of the observed autocorrelation and suggest ways to deal with it. In this regard, the Markov chain model was at least partly successful: It predicts that autocorrelation should be present, and even predicts that the form of the autocorrelation should be first order.[28] We have confirmed that both the F/A–18 data shown in figure 9 and similar data for F–14s do indeed exhibit first order autocorrelation. Still, the more conventional modeling approaches have nothing to say about whether autocorrelation should be expected, and so the Markov chain model can claim to be able to explain at least that characteristic of the data in a way that a more conventional model cannot.

Full technical details of the model and associated estimation procedures are in appendix B. In the following sections, we provide a nontechnical description of the model, discuss some of its theoretical implications, and present some preliminary empirical results based on data for F/A–18s and F–14s.

Model structure

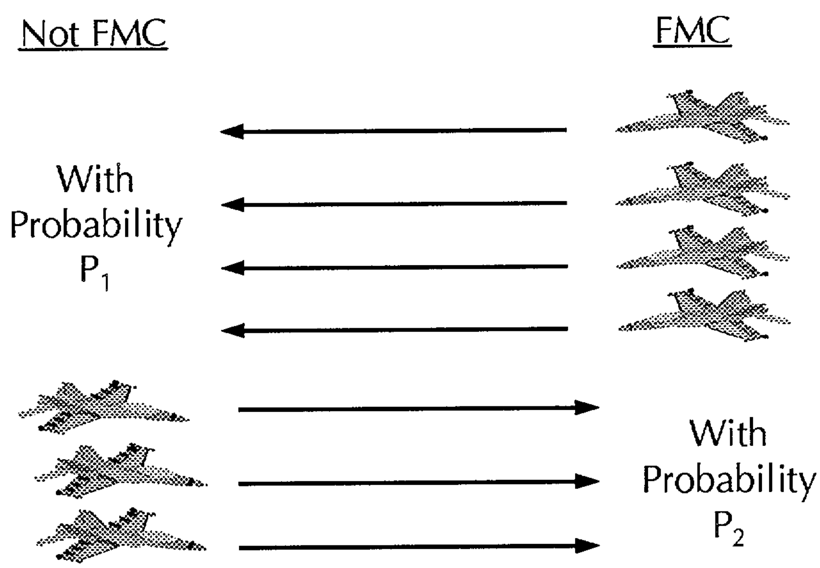
The basic idea behind the Markov chain model is that we consider what is going on at the level of the individual aircraft. Although we have no data on individual planes, we can nonetheless think about a model of aircraft entry into and exit from the FMC state. We propose a simple mechanism that characterizes the probability of any given plane's moving into or out of FMC, and then deduce from that what the behavior of the overall fleet FMC rate should look like.

Figure 11 illustrates the essence of the model. We suppose that FMC is a discrete state and that each plane, if it is currently in that state, has a fixed and independent probability P_1 of leaving that state during

28. First order autocorrelation means that any given observation is correlated with its immediate predecessor only; if you know the value of the immediate predecessor, no other observation further back in time has any additional predictive power.

the observation period. If a plane is currently not FMC, it has a probability P_2 of reentering the FMC state during that period. Given this basic structure, it is possible to derive the probability distribution for the state of the system at any time given that you know the state at some initial time. Full details of that derivation and the conclusions that can be drawn from it are in appendix B.

Figure 11. Diagrammatic representation of Markov chain model

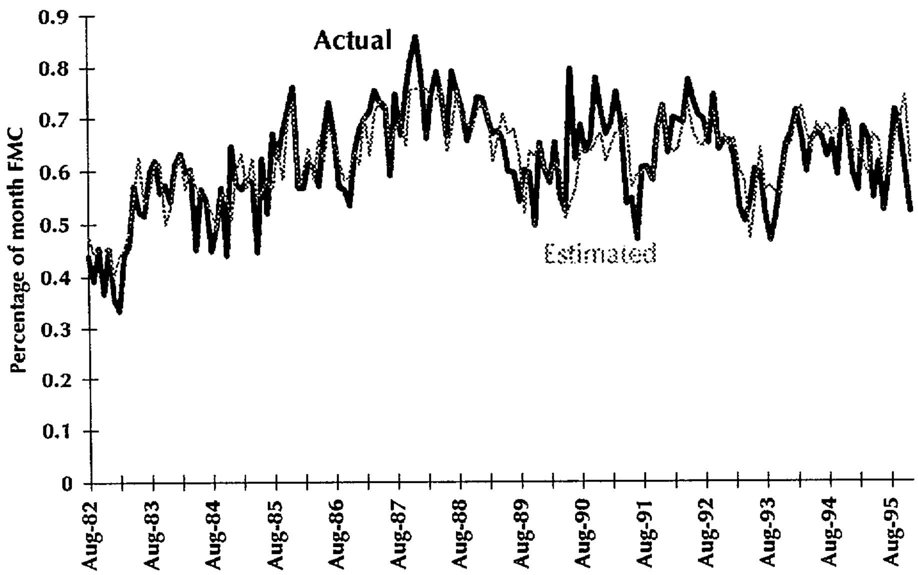


Development of this model is not yet complete, and the analysis presented here should be considered exploratory. However, we have done some empirical analysis of a subset of the available data, and the results so far suggest that the Markov chain model holds promise. That optimism needs to be qualified, though, because estimation of this model has presented some difficulties that are a bit out of the ordinary.

Preliminary estimation results—the impact of drivers on movements in and out of FMC

With the above caveats, we now report results from applying the Markov chain model to F–14 data. Figure 12 shows actual F–14 FMC rates (Navy-wide) graphed with corresponding predicted values. We produced the latter by applying the Markov model to a panel of individual squadron observations and averaging the resulting predicted values. Note that the results in figure 12 are for deployed squadrons: We estimated deployed and nondeployed regressions separately. This was done in the belief that explanatory variables may have different effects depending on a squadron's deployment status; the results described below would appear to bear that out.

Figure 12. Actual and predicted FMC rates for deployed Navy F–14s



Actual and predicted FMC rates appear to correspond reasonably well. Note, however, that the predicted values are dependent on the previous month's actual values. Such relationships tend to give the appearance of a close correspondence, so we do not ascribe great importance to this as a goodness-of-fit indicator. Of greater interest are the estimated effects of explanatory variables. Generally, we found that measures of personnel readiness, supply, and the number of sorties were strongly associated with FMC rates. Of particular interest are how some of these determinants affected the probability of leaving FMC versus the probability of returning to FMC. Table 6 shows how the significant variables affect the FMC rate.

Table 6. The impact of drivers on the FMC rate

	Resulting changes in the FMC rate	
Significant drivers	Deployed	Nondeployed
Sorties per aircraft	-0.61	1.14
Training SORTS (days in C1)		0.30
Percentage of requests for repairables filled in less than a day		1.2
Officer turnover (3 months)	0.025	
Enlisted personnel quality	[a]	[a]

a. The impact of personnel quality is described in the following section.

We found that the number of sorties per aircraft was significant for both deployed and nondeployed F–14s. As before, the effect was found to be of opposite signs according to whether the squadron was deployed or not, with additional sorties having a detrimental impact for deployed squadrons and a beneficial effect for nondeployed ones. This also appears to validate the notion that the deployed and nondeployed states are qualitatively different.

Training readiness had a significant, positive effect on nondeployed FMC rates. A simple explanation for this would be that better trained pilots are less likely to damage their aircraft. However, caution is warranted when interpreting this result. The direction of causality could run in the other direction; more training may occur in squadrons that have healthier aircraft. Further investigation is needed to deal with the possible simultaneity in this relationship.

The number of repairables on hand is a significant positive contributor to FMC rates for nondeployed F–14s. This variable is particularly interesting in the way it enters into the model, because we found that it mattered mainly on the breakdown side rather than on the repair side (that is, P_1 rather than P_2). The implication is that having spares readily available seems to help prevent planes from dropping out of the FMC state, perhaps because they're repaired so quickly that there is no need to report them as not FMC.

We also found that officer turnover was significant for the deployed F–14s. The impact was adverse and was felt mainly on the breakdown side. Personnel quality, as measured by the personnel quality index, was found to be significant in the deployed regression. It was important on the repair side and had a beneficial effect, which is what we would expect.[29]

Possible extensions

Much more can be done both to firm up the work already developed and to use the model with other aircraft TMS and other applications. There are still unresolved methodological issues associated with numerical estimation algorithms, goodness-of-fit measures, and hypothesis tests. There are also potential uses for this approach in short-run forecasting. To this end, we are looking at the possibility of verifying some of the forecasts by working with other datasets that may allow us to independently determine the entry and exit rates to and from FMC status. Still, the difficulties inherent in this line of attack should not be underestimated. We have reason to hope this approach will be useful, but we cannot prove at this point that it is superior to a more conventional modeling approach.

[The absence of] failures and the FMC rate

Our empirical findings did not support our prior expectation about how the number of flight hours between failures affect the FMC rate. We offer some explanation. First, the number of flight hours between failures are merely a standardized way of counting the number of reported failures. Many reported failures are not mission degrading. Because so many are rather minor and can be repaired quickly, they

29. We tried other variables in these equations, but the variables were not significant. They are listed in appendix B.

never affect the FMC rate. At this point, we have found no way to sort serious failures from trivial ones. This may be one reason why we are unable to link failures to the FMC rate. A second reason may be that higher failures are accompanied by larger stocks of spare parts, a theory that is partially supported by our repairables equation.

Training

Again we turn to SORTS for our readiness proxy. In this section, we explain how determinants are linked to a squadron's SORTS training score.

SORTS training reflects the status of each squadron's set of training objectives. Squadrons aim to meet all the training requirements as dictated by their training matrix.[30] We want to show what things affect the ability to meet these goals. Because training in any one specific task does not expire daily or even monthly, the current level of determinants (like flight hours) cannot explain the current levels of training. We need to look at some aggregation of these drivers. Often training in any one event is good anywhere from 3 months to nearly 2 years. Since most of the events seem to expire well within the first year, we considered aggregating training drivers over a 6-month period to be a reasonable approximation.

We expect the quantity of flight hours accrued over the last 6 months to have a significant effect on the amount of training the unit has had the opportunity to accomplish. We also expect that the equipment readiness would act as a constraint—only equipment fully able to perform its missions is able to train to those missions. In light of this theory, we look at the effect of a 6-month accrual to FMC flight hours (the FMC rate times flight hours).

We also expect a squadron's deployment cycle and deployment history to affect its training readiness. Specifically, we expect squadrons that are deployed to be more ready than those that are not. Nondeployed squadrons should show progress in completing their training

30. Note that SORTS reflects the act of training for a specific task, not the squadron's proficiency at that task.

matrix as the next deployment draws near. On the other hand, we also expect that back-to-back deployments will have an adverse impact on training opportunities and ultimately on training readiness. Furthermore, operating in areas that restrict squadron training opportunities (such as the Adriatic, Haiti, and Somalia) should affect readiness because of limited airspace, training ranges, etc. On the one hand, we expect that the reduction in training opportunities would jeopardize training readiness, especially if the unit was in that area for an extended period of time. On the other hand, only the most ready units would be deployed to those areas and, once there, they would be given the highest priority for resources needed to maintain their readiness status.

Another factor that may influence the ability of units to accomplish their training is their current stock of ordnance, specifically training ordnance. We could not find historical data on individual squadron or airwing inventories over time, so we used a rather crude proxy—the Navy's inventory of missiles relative to platform requirements. This measure reflects inventory relative to design requirements each year. We expect that more missiles relative to requirements is associated with higher training readiness.

The number of officers should also be positively associated with training readiness. More officers should mean more opportunities (flexibility) for the squadron to meet its requirements. Given that more officers implies a tighter distribution around the average (the law of large numbers), we are assuming that more officers near the average implies more officers trained to C1 standards.[31]

Finally, we expect that the F/A-18s are harder to train than the other types of fighter and attack platforms. F/A-18s have half the number of officers required to train for twice the missions.

31. We are also assuming that the squadron has enough money to support these officers.

Estimation method

We initially tried to estimate a model where determinants were used to explain the duration or number of days squadrons spend in C1 and the number of days they spend not C1. The idea here is that different factors were important for determining the length of C1 and not C1 durations. One might envision a scenario where it takes different talents to move a unit from unsatisfactory to satisfactory performance than it takes to maintain a unit at a satisfactory level.

Our results, while believable, did not support this hypothesis. They revealed that whatever prolongs a non-C1 period will prolong a C1 period if it is reversed. In this case, it is better to model training readiness as if it were a complete process, so we did.

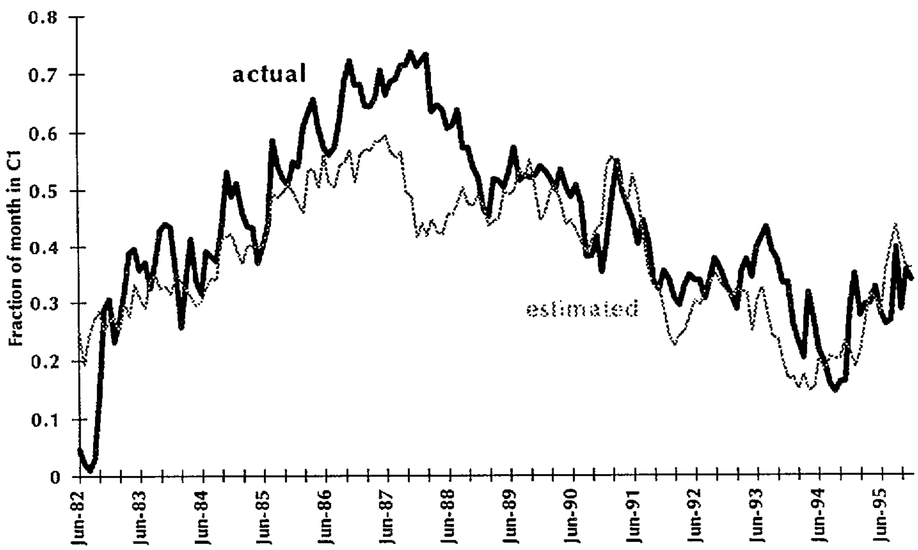
We used regression analysis to explain the percentage of time that a squadron is C1 for training at any given month.[32] Equation 6 represents our model.

Training = *T(deployment indicator, historical deployements, FMC flight hours, carrier equipment readiness, officer turnover, the number of officers, TMS indicators)* (6)

We find that the coefficients as a group are highly significant in explaining the odds of being C1 for training. Figure 13 shows how our predicted values match the actual values over time. Although we underestimated training readiness during the mid 1980s, we have been close to the actual levels since then. In fact, for some periods, it appears that our estimates move just before the actuals; our predictive values are acting as leading indicators—perhaps because we use 6-month stocks of drivers rather than a longer period, say 8 months.

32. We estimate training readiness using a logistic equation. Details are in appendix B.

Figure 13. Actual versus estimated training readiness



Estimation results—how drivers affect training readiness

Our results generally support our prior assumptions about the drivers of readiness. Table 7 shows the size of the most significant drivers' effects.

We did not find that the quality of enlisted crews had a significant direct effect on time spent in C1 for training. Recall that we did find a significant relationship between personnel quality and FMC rates, and we do see that FMC rates influence training readiness. Therefore, while personnel quality may not directly affect training, there is an indirect effect through the FMC rates.

The relationship between flight hours and training readiness is nearly definitional—flight hours are a necessary component of training readiness. However, all flight hours do not contribute equally to training. Our model supports our claim that FMC flight hours, specifically the 6-month accrual of FMC flight hours, are important to training readiness. A 10-percent increase in the amount of FMC flight hours accrued over a 6-month period is associated with nearly 2 additional days in C1 for training every month.

Table 7. The impact of drivers on training readiness

Significant drivers	Resulting change in training readiness
Deployed relative to not deployed	19%
A7–E relative to A6–E	24%
F–14A relative to A6–E	25%
F–14B relative to A6–E	-2%
F–14D relative to A6–E	-11%
F–4S relative to A6–E	53%
F/A–18A relative to A6–E	-33%
F/A–18C relative to A6–E	-29%
A 10% increase in the average number of days the carrier is C1 for equipment	1%
A 10% increase in the number of FMC flight hours accrued over the last 6 months	6%
A 10% increase in 6-month officer turnover	-1%
A 10% increase in the number of months deployed over the last 6 months	-1%

The equipment readiness of the carrier is also relevant to the training readiness of the squadron. Lowering the carrier's equipment readiness by 10 percent could cause the squadron's readiness to fall by nearly 1 percent.

We looked for a link between the Navy's stock of missiles and VF and VFA squadron readiness and found one. The relationship is positive indicating that higher inventories are associated with higher training readiness.[33]

33. We only had data for missile inventories. Therefore, we had to run a separate regression excluding VA squadrons to test for a link between ordnance and training. Coefficients for these results are available from the authors.

The 6-month turnover of the squadron's officers is associated with a statistically significant decrease in training readiness. A 10-percent increase in turnover will reduce time in C1 for training nearly 1 percent.

Probably the most significant drivers of aircraft training readiness are the deployment status and the TMS. Once a squadron is deployed, we find that it spends 19 percent more time in C1 for training.[34] Some of the TMS indicators have even larger effects. The most notable is the F/A–18C. It appears that training in the F/A–18 is much harder than training in the other platforms, which supports our prior theory that doubling the mission using half the crew presents a training challenge for this aircraft.

34. We could not find a link between squadrons operating in restricted zones and training readiness. It may be that our deployment variables have picked up much of this effect.

Personnel quality and squadron readiness

We've found that the personnel quality of a squadron plays an important role in the readiness of that squadron. That personnel quality is important seems rather obvious; the real question is how important is it? The problem is that personnel quality is a concept—an indicator we all understand yet have no measure for. To provide a clear, tractable measure of such an important indicator, we have computed an index of the quality of every squadron's personnel. With this measure in hand, we are free to return to our equations to explore how a change in personnel quality is associated with changes in our measures of readiness.

The personnel quality index

Our index measures information about the experience and intellectual performance of squadron personnel every month from 1982 through 1995. Note that our measure of quality refers not to recruits, but to people who are actually assigned to a unit at any month in our sample. It is the stock (the working Navy) rather than the flow (the recruits) that are relevant to today's readiness.

In computing our index, we collected data on five variables that we believe reflect the quality of squadron personnel. There are probably other measures that are entirely relevant to this measure and could be added at some point. Our challenge was to find measures that are consistently measured over time and for individual units. We used the following:

- The percentage of the squadron with a high school degree
- The percentage testing in the upper mental group (categories I, II, and IIIa) on the Armed Forces Qualification Test (AFQT)
- The percentage of the squadron demoted that month

- The percentage of the squadron who were promoted to E5 or higher with less than 4 years of experience
- The average length of service of the squadron's enlisted personnel.

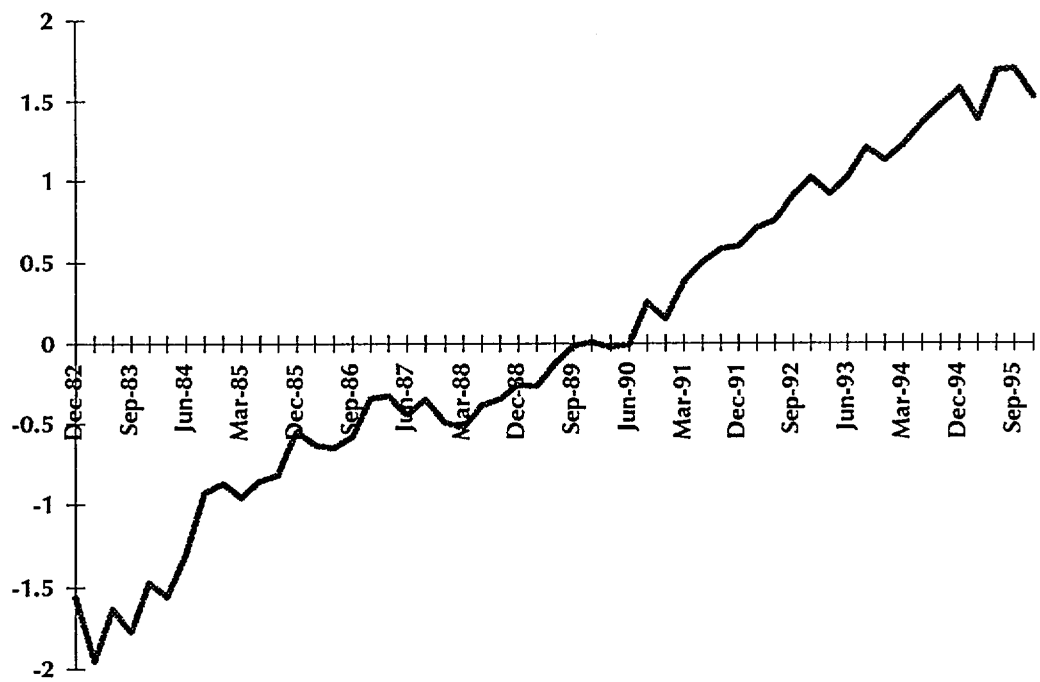
We combined these five measures into one index using a mathematical technique that resembles a weighted average. The difference is that our weights are not arbitrarily chosen; they are chosen based on their ability to capture as much information from the data as possible.[35] The result is one variable that captures most of the information available from the five individual variables. When we computed the index monthly for individual units, we had a new measure of personnel quality for use as a variable in the regression analysis presented in the preceding section.

The new measure is in z-score or standard deviation units. In other words, the units are standardized so that across the data, the average value of the index is 0 and the standard deviation is one. While these units are informative to some, most cannot get a feel for how to interpret an index value of, for example, -1.3. One way to handle this problem is to take advantage of the historical nature of the data and use historical benchmarks (the hollow period of the late 1970s and early 1980s, for example) to add meaning to the index values.

Figure 14 shows how our squadron personnel quality index looks when computed for average squadrons over time. We can easily see that our index captures the personnel quality problems of the 1970s and early 1980s nicely. Since then, the data show a dramatic increase in quality. In fact, the final 1995 value appears to be nearly four standard deviations higher than it was in 1982. The average value of personnel quality over this time series was in 1989. This is startling if you consider that we fought Desert Storm with nearly that quality and we are now about 1.5 deviations better than that position as of the end of 1995.

35. See appendix C and Junor and Oi, 1996, for technical details.

Figure 14. The personnel quality index for active duty squadron personnel



What is the effect of high-quality sailors on readiness?

Now that we have established some meaning for our measurement of personnel quality, we return our attention to assessing how a change in quality affects readiness. We know from our regression results that personnel quality has a significant impact on personnel readiness, equipment readiness, and the proportion of local repairs. We also know that equipment readiness affects training readiness. With these equations in hand, we have the tools we need to measure the impact.

To illustrate how squadron quality affects readiness, we will run an experiment in which we ask ourselves: How different would 1995 squadron readiness look if the high-quality 1995 crews were replaced with lower quality 1982 crews? We acknowledge that this is a huge change in quality, one not likely to happen overnight. However, we believe that it provides an interesting illustration of how quality affects readiness.[36]

To answer this question, we will use our equations to calculate an estimated level of readiness using 1995 levels of everything—supplies, deployments, manning, flight hours—everything. We will put these estimates of personnel, equipment, and training readiness aside. Then we will go back and mathematically "rip out" the 1995 values of squadron quality and replace them with the lower 1982 values, leaving everything else at the 1995 levels. We will then recalculate our readiness estimates and compare the new 1982 quality estimates with the 1995 estimates we were setting aside. The only thing that changed in these calculations was the quality of the squadron personnel, so any difference between these numbers can be entirely attributed to this very large difference in quality. Figure 14 illustrates what we saw when we compared the two sets of numbers. The figure expresses the changes as percentage changes from the original 1995 value. Table 8 shows the actual change relative to the average value over the entire time period.

The first thing we noticed was the large impact on the percentage of time squadrons are C1 for personnel. The change in quality from the 1995 level to the 1982 level caused time in C1 to fall from 81 percent to only 12 percent—85 percent as a percentage of the original value. Referring to table 8, we see that the reduction in quality brought personnel readiness far below the 13-year average.

Remember from our last section that we did not find a relationship between *squadron* personnel and the proportion of AIMD reported repairs that were done on site, but we found that the *ship's* enlisted quality did matter. This was not surprising since the ship's AIMD personnel probably have the most influence over those repairs. Conducting the exact same experiment using ship's personnel quality, we see

36. Again, we acknowledge that this is a relatively unrealistic change in personnel quality, one not likely to occur in a short period of time. We offer it only as an illustration. Certainly, smaller changes in quality can be evaluated.

Figure 15. The impact of a change in personnel quality (from 1982 to 1995 levels) on readiness measures

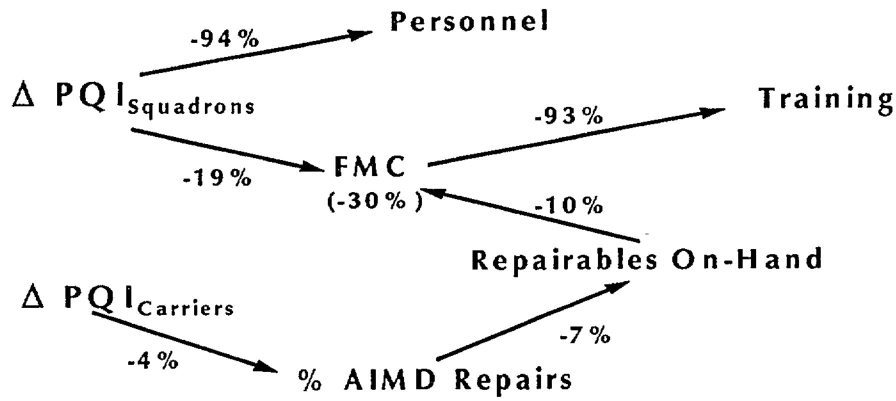


Table 8. Changes in readiness relative to historical values

Variable	Predicted 1995 readiness	Predicted 1995 readiness after substituting 1982 personnel	Historical average (1982 through 1995)
Personnel SORTS	81%	13%	45%
FMC rates	68%	47%	60%
Training SORTS	25%	11%	74%
AIMD repairs	56%	54%	57%

that the change in quality caused a modest 4-percent drop in the number of onsite repairs.

The impact of the change in squadron enlisted personnel also had the effect of decreasing the FMC rate from 68 percent to 47 percent.[37] This 31-percent change pulled the FMC rate below its historical average. Remember, this change left the values of all other drivers constant at the 1995 levels. Therefore, despite the fact that the Navy

37. To do this calculation, we used our first method of estimating FMC.

had more dependable aircraft in 1995, the reduction in quality is still associated with reducing their ability to maintain the aircraft. The reduction in FMC rates also had a sizeable impact on training readiness. As a result of this experiment, we see that training readiness fell by 57 percent of its 1995 value.

Concluding remarks

The purpose of our research is to increase our understanding of readiness and its causes. If we clearly understand what readiness is and why it changes, we are in a better position to manage it. To develop this type of understanding, we analyzed relationships in a rich historical database. What follows describes some of our most notable findings and how they can be used to manage readiness.

- *We have found an empirical link between measures of readiness (such as SORTS and FMC rates) and their determinants.* This link:
  - — *Highlights which determinants are most important to readiness.* Once such determinants are identified, the Navy is in a much better position to protect them or at least monitor their movements. For example, we found that personnel quality remains one of the most important determinants of readiness. We also find that FMC flight hours, not just flight hours, are important to training readiness. Often we hear that flight hour funding needs to be maintained in order to protect training. But we've also found that making sure these aircraft can do their missions is also extremely important.
  - — *Highlights possible consequences of changing the level of these determinants.* Generally, we are able to show what happens to readiness given a change in one of the determinants. Following from above, we show that changes in personnel quality and FMC flight hours, for example, have a significant impact on readiness. If the services cannot afford to maintain the level of these determinants, at least they can use this type of analysis to understand the risk of lowering them.

- Our modeling also suggests that:
  - *Different skills are required to move a squadron to FMC than to keep a squadron in FMC status.* This type of knowledge again helps us understand how determinants work to affect readiness.
  - *There are still broad historical sweeps that we don't understand.* For about a 5-year period in the early to mid 1980s, we are consistently underestimating readiness. In fact, the error terms for flight hours between failure and training SORTS appear to be correlated over this time period. This suggests that there is some higher level variable, possibly one measuring funding, that is driving several of our readiness measures.
- *We evaluated the history of readiness, its current trend, and what is driving our current level of readiness.* We're not sure we can forecast what readiness will be in the future and therefore prevent damage to readiness, but we can at least understand what is causing readiness to degrade now and address these issues before things get worse.
- *Most readiness indicators appear to have leveled off at fairly high values; however, training appears to be a potential weak spot.* Unfortunately, the SORTS reason code data cite "other" as one of the most popular reasons for training deficiencies—which is of no value analytically. By reconstructing our model estimates and comparing them to levels of the most important drivers, we can use this analysis to find out more about what has been happening with training readiness over the last few years.
- *Indexing or summarizing competing measures of readiness are good tools for evaluating historical trends.* We've extended our application of indexing from simply measuring personnel quality to finding the best single measure of equipment condition. In the latter case, we found that the FMC rate summarizes the information given by several other measures of equipment condition such as mission capable rates, partially mission capable rates, and equipment SORTS. We also found that equipment SORTS measures a second dimension of equipment readiness and should not be considered a substitute for FMC.

- *Our results suggest that technology is related to readiness.* We found that the TMS of aircraft are important determinants of readiness. This is an important finding because we have a good idea of what our TMS profile will be over the next several years. It also implies that it is possible to affect future readiness by placing a priority on designing reliable systems. The F/A–18 is the prime example.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A: Choosing the best measure of equipment readiness

Given the number of variables that measure the material condition of aircraft, we wanted to determine which indicator or combination of indicators best describes the available information. We collected data on the following variables monthly from 1982 through 1995 for each squadron:

- The percentage of aircraft that are partially mission capable due to maintenance
- The percentage of aircraft that are partially mission capable due to supply
- The percentage of aircraft that are not mission capable due to maintenance
- The percentage of aircraft that are not mission capable due to supply
- The percentage of aircraft that are mission capable
- The percentage of aircraft that are fully mission capable
- The percentage of time a squadron is C1 for equipment in SORTS.

Our approach gathers these competing variables and uses a mathematical technique called principal component analysis to compute something like a weighted average of the original variables.[38] The process computes as many index values as there are original variables.

38. This is the same approach we used to compute the personnel quality index (PQI) for aircraft squadron enlisted personnel used in this research or for the surface combatant used in Junor, 1996.

Each additional index explains successively less of the information available.

Our first index of material condition (MCI) explains about 76 percent of the available information. The weights are listed in table 9.

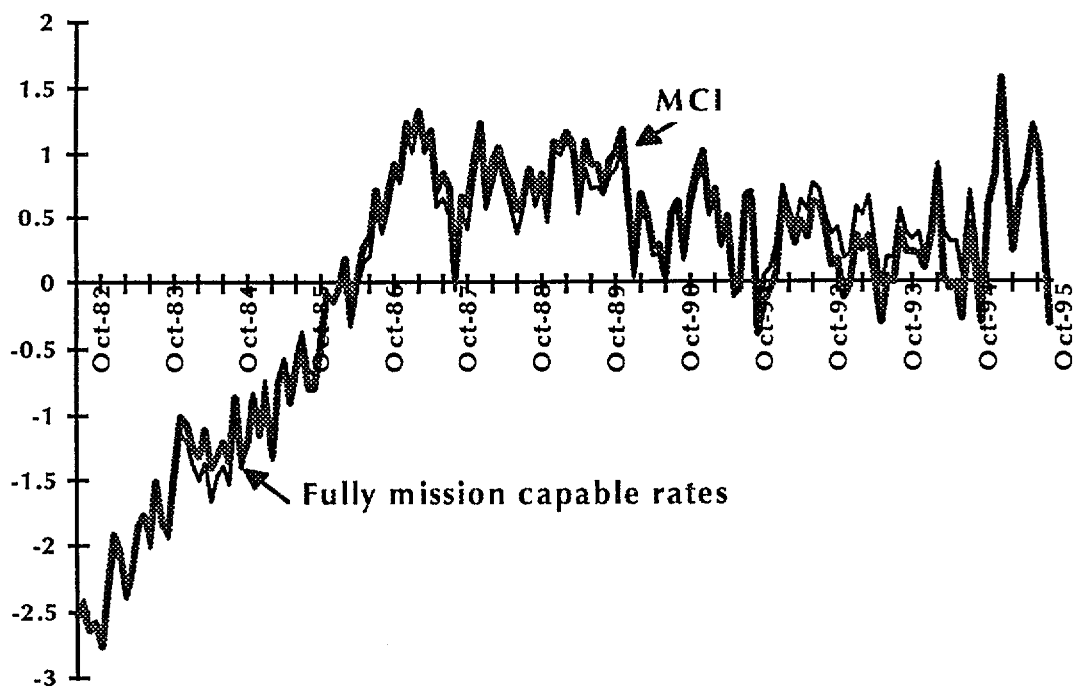
Table 9. MCI weights

Variable	MCI (1) weights	MCI (2) weights
Partially mission capable due to maintenance	-.37	.45
Partial mission capable due to supply	-.37	.40
Mission capable	.42	.15
Fully mission capable	.43	-.14
Not mission capable due to maintenance	-.40	-.03
Not mission capable due to supply	-.37	-.31
Equipment SORTS	.26	.70

When we graphed the data (see figure 16), we saw that the index was better than 99 percent correlated with the fully mission capable (FMC) rate. Given this result, we concluded that there was no advantage to using this index. FMC gives the best summary measure of aircraft equipment condition.

We did notice, however, that this index weighted equipment SORTS significantly lower than any of the other measures. Upon further inspection, we found that the second index was almost entirely weighted on equipment SORTS. The conclusion here is that equipment SORTS measures an entirely different dimension of aircraft equipment condition than does the FMC rate.

Figure 16. The material condition index (MCI) relative to FMC



THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B: Technical notes on estimation methods and marginal analysis

Estimating dependent variables that are in the form of percentages (like percent of time in C1 or FMC) cannot be done using a simple, linear regression technique like ordinary least squares (OLS). The biggest problem is that the dependent variable is bound by 0 and 1 and, unless this issue is resolved, a linear technique will estimate values outside these bounds.[39] For this reason, we have chosen alternative estimation methods such as a logistic approach or a Markov process (using nonlinear least squares) to estimate personnel readiness, training readiness, and FMC rates. The problem with these methods is that the results are often difficult to interpret. For example, the estimated coefficients are not the marginal values that we would get had we used OLS. There are also some problems in evaluating the standard errors and testing for significance.[40]

In this section, we explain the details behind our estimation techniques and the calculation of the marginal values that appear in the text.

39. OLS, in fact, yields biased estimates in these cases. Greene (1993) provides a detailed explanation of limited dependent variables and the problems they pose in estimation.

40. The flight hours between failures is reasonably well behaved, and we estimated them using OLS.

Estimating personnel readiness, training readiness, FMC, repairables on hand, and the proportion of AIMD repairs using a log-odds approach

One approach to estimating bounded variables is to transform the data using a logistic or log-odds approach. Suppose we have data on the amount of time squadrons spend in C1 for training every month (denoted *T1*). We can transform this variable into one that reflects the log of the odds of being *T1* for the entire month:[41]

$$oddsT1 = [(T1)/(1-T1)] \; . \tag{7}$$

OddsT1—also called the logit of *T1*—is now linear and can be estimated using OLS:[42]

$$oddsT1 = \beta\chi \; . \tag{8}$$

Tables 10 through 14 give the reported coefficients for each of our log-odds equations.

41. To do this calculation, we had to adjust the data so that when $T1 = 0$, we recoded it to $T1 = 0.0001$, and when $T1 = 1$, we recoded it to $T1 = .9999$.

42. This equation is heteroscedastic when applied to FMC and the two SORTS equations. For this reason, we tried estimations using weighted least squares. The heteroscedasticity enters in the FMC equation because FMC is calculated for each individual aircraft, although we see only a grouped or proportional version of these data. In other words, our FMC rate is actually the proportion of some number of airplanes n that were FMC. Each airplane has the same x_i. The error variance for every observation i is $1/(n_i \, FMC_i(1\text{-}FMC_i))$ where n is the group size. The heteroscedasticity enters into the SORTS equation because we (potentially) observe daily SORTS scores yet are estimating percentage of reporting days that are C1. Since the number of reporting days differ among observations, we used a similar weighting method to that described above for FMC. Instead of adjusting for the number of aircraft, we adjust here for the number of reporting days. Since the weighted version of the results was nearly identical to the original unweighted version, we report the original.

Table 10. Reported coefficients for the logit of time in C1 for personnel

Variable	Coefficient	Significance[a]
Intercept	-6.8	**
A–7E	7.8	**
F–4S	1.3	*
F–14A	0.9	**
F–14B	2.7	**
F–14D	1.6	**
F/A–18A	7.9	**
F/A–18C	5.6	**
Deployment indicator	1.2	**
Enlisted personnel quality	3.2	**
Enlisted 3-month turnover	-0.1	**
Officer 3-month turnover	-0.04	**
Enlisted manning relative to requirements	0.04	
Number of observations	9012	
R^2	.28	

a. A * means this variable is significant within a 90% confidence interval; a** means this variable is significant within a 95% confidence interval.

Marginal effects

The reported coefficients are not the marginal effects of a change in a determinant of *T1*. Our objective is to relate a change in a determinant to the percentage of time ready *T1* rather than the logit. Since *T1* is nonlinear, any marginal values calculated for it will vary depending on where the derivative is taken. We found that the easiest way to understand the size of the impact of one determinant on our readiness measures was to evaluate a general elasticity taken over all observations. Specifically, we derived an estimate of *oddsT1* at all observations (for every squadron and every month) both before and after a 10-percent change in one of the determinants or x values. We used the following to transform *oddsT1* back to *T1*:

$$T1 = e^{\widehat{\beta \chi}} / (1 - e^{\widehat{\beta \chi}}) \,. \qquad (9)$$

Comparing average *T1* before a change with average *T1* after a change gives us a reasonable idea of how a change in x affects

Table 11. Reported coefficients for the logit of time squadrons are FMC

Variable	Coefficient	Significance[a]
Intercept	-0.1	**
A–7E	-0.1	**
F–14A	-0.01	
F–14B	-0.02	
F–14D	-0.1	
F/A–18A	0.06	**
F/A–18C	0.06	**
Deployment indicator	0.6	**
Months since last deployed	0.004	**
Repairables on hand	0.002	**
Consumables on hand	0.001	*
Sorties per aircraft	0.01	**
Sorties per aircraft—deployed	-0.03	**
Carrier equipment readiness	0.01	**
Enlisted 3-month turnover	-0.01	**
1/(4+personnel quality index)^{2}[b]	-2.4	**
Lagged FMC	0.7	**
Number of observations	8422	
R^2	.53	

a. A * means this variable is significant within a 90% confidence interval; a ** means this variable is significant within a 95% confidence interval.

b. We used this transformation to prevent the effect of personnel quality from turning negative.

readiness. We used the same calculation for our estimates of the percentage of AIMD repairs, the percentage of time in C1 for personnel, and the percentage of time in FMC (our first method of estimating FMC). However, the FMC equation included a lagged value of the dependent variable as a determinant, which requires a few more calculations.

Let Y equal the "log-odds of being FMC":

$$\Upsilon = \log[(T1)/(1-T1)] \ . \tag{10}$$

We see that

$$\Upsilon = \beta\chi + \alpha(lag\Upsilon) \ . \tag{11}$$

Table 12. Reported coefficients for the logit of time in C1 for training

Variable	Coefficient	Significance[a]
Intercept	-8.0	**
A–7E	2.5	**
F–4S	4.8	**
F–14A	2.2	**
F–14B	-0.2	
F–14D	-5.6	**
F/A–18A	-4.1	**
F/A–18C	-3.8	**
Deployment indicator	2.1	**
Accrued FMC flight hours, 6 months	0.0001	**
Carrier equipment readiness	0.03	**
6-month deployment history	-0.6	**
Average number of officers over 6 months	0.4	**
Officer 6-month turnover	-0.04	**
Observations	8705	
R^2	0.18	

a. A ** means this variable is significant within a 95% confidence interval.

Table 13. Percentage of AIMD reported repairs done at the AIMD

Variables	Coefficients	Significance[a]
Intercept	0.04	
A–7E	-0.01	
F–4N	-0.2	**
F–4S	0.01	
F–14A	0.1	**
F–14B	0.1	**
F–14D	-0.1	**
F/A–18A	-0.1	**
F/A–18C	-0.1	**
Deployment indicator	-0.1	**
Years of service of TMS	0.01	**
Number of days squadron is C1 for personnel	0.0002	
Enlisted personnel quality for ship	0.04	**
Number of failures	0.0001	**
Consumables on hand	0.001	**
Observations	8630	
R^2	0.14	

a. A ** means this variable is significant within a 95% confidence interval.

Table 14. Percentage of requests for repairables met within one day

Variables	Coefficients	Significance[a]
Intercept	-0.1	
A–7E	0.1	
F–4N	-3.3	**
F–4S	-0.3	
F–14A	-0.9	**
F–14B	-0.01	
F–14D	-0.6	*
F/A–18A	-0.7	**
F/A–18C	-0.7	**
Deployment indicator	0.7	**
Months since last deployed	0.02	**
Percentage of items repaired at the AIMD	2.9	**
Number of failures	0.0001	
Observations	8611	
R^2	0.04	

a. A ** means that this variable is significant within a 95% confidence interval.

Solving for the marginal values of the percentage of time in FMC at the steady state requires the following calculations:

$$steady(\Upsilon) = ((\widehat{\Upsilon} - \alpha(lag\Upsilon)))/(1-\alpha) = (\beta\chi)/(1-\alpha), \quad (12)$$

$$steady(FMC) = e^{(\beta\chi)/(1-\alpha)}/(1+e^{(\beta\chi)/(1-\alpha)}) . \quad (13)$$

Steady(FMC) is evaluated at the average of all observations before and after a 10-percent increase in one of the x values as before.

The deployment status and TMS variables are dummy variables—variables that only take on values of 0 or 1. Since it does not make sense to increase this type of variable by 10 percent, we used the equations above to estimated readiness for all deployed squadrons. Then, we pretended these squadrons were now nondeployed by changing the value of the deployment variabe from 1 to 0 and recalculated readiness. The only variable we changed was the value of the deployed variable. The difference should reflect the impact that being deployed had on readiness. We calculated the impact of the TMS variables the same way.

Estimating flight hours between failures using OLS

The number of flight hours between failures calculated per squadron per month is continuous and unbounded. We estimate it using OLS. Table 15 lists the estimated coefficients.

Table 15. Reported coefficients for flight hours between failures

Variables	Coefficients	Significance[a]
Intercept	0.04	
A–7E	0.2	**
F–4N	0.2	**
F–4S	0.2	**
F–14A	0.03	*
F–14B	-0.1	
F–14D	0.4	**
F/A–18A	0.8	**
F/A–18C	1.1	**
Deployment indicator	0.2	**
Months since last deployment	0.01	**
Flight hours per aircraft	0.01	**
Sorties per aircraft	0.001	**
Sorties per aircraft—deployed	-0.01	**
Number of aircraft	0.01	**
Observations	8644	
R^2	.44	

a. A * means this variable is significant within a 90% confidence interval; a ** means this variable is significant within a 95% confidence interval.

In this case, the reported coefficients are the marginal values. For completeness, however, we calculated elasticities for this equation in the same general manner as the other equations—we used the estimated coefficients to calculate flight hours between failures for every squadron for every month and compared the average values before and after a 10-percent change in a determinant. Because there are no nonlinearities are involved, the value of the marginals taken at the averages and the average of the marginals taken at the individual observations should be the same.

Estimating FMC using a Markov process

Introduction

Here we present details of the structure, derivation, and application of the Markov chain model. In the first section of the appendix, we described the model without explanatory variables; we did this to make the underlying concepts and intuition as clear as possible. The next section describes the way in which explanatory variables can be incorporated into the model and explains why we chose the particular functional form that we did. In later sections, we take up heteroscedasticity, the treatment of varying numbers of aircraft, and the estimation methods and variables used.

Basic model

We construct the basic model by supposing that there are two complementary probabilistic mechanisms working on individual aircraft. A plane is assumed to be in one of two discrete states: FMC or not FMC. A plane that is currently FMC is assumed to have a probability P_1 of being not FMC next period, and therefore a probability of $1 - P_1$ of still being FMC in the next period. A plane that is currently not FMC is assumed to have some probability P_2 of returning to FMC next period—and, of course, a probability of $1 - P_2$ of remaining not FMC. In general, P_1 and P_2 are not directly related to each other. We will assume for the moment that all aircraft are the same and that the outcome for one plane is independent of what happens to any other plane. Now, take the total number of planes to be fixed as N. Then the total number of planes that are FMC next period is the sum of two binomial random variables as follows:

$$FMC_{t+1} = Bi(FMC_t, 1 - P_1) + Bi(N - FMC_t, P_2) \quad . \tag{14}$$

Since $E[Bi(x, P)] = xP$, the expected value of the FMC rate at time $t+1$, given the FMC rate at time t, is

$$E\langle FMC_{t+1} | FMC_t \rangle = (1 - P_1) FMC_t + P_2 (N - FMC_t) \quad . \tag{15}$$

It is then straightforward to solve for a stationary point in the process by setting $E\langle FMC_{t+1}|FMC_t\rangle = FMC_t$. The result is

$$FMC_{ss} = \frac{NP_2}{P_1 + P_2} \quad . \tag{16}$$

We use the "*ss*" subscript to evoke the idea of a steady state; note, however that FMC_{ss} is not—strictly speaking—a steady state. If $FMC_t = FMC_{ss}$, it does not in general follow that $FMC_{t+1} = FMC_{ss}$ because of the random fluctuation associated with the process.[43]

Next, define $\alpha \equiv P_1 + P_2 - 1$. We can then manipulate equation 15 to read

$$E\langle FMC_{t+1}|FMC_t\rangle = FMC_{ss} - \alpha\,(FMC_t - FMC_{ss}) \quad . \tag{17}$$

Conveniently, this is a standard form for a first order autoregressive process and is denoted by AR(1). To estimate this equation, however, we must consider the nature of an associated error term. This is not unduly difficult, although one complication does arise. The good news is that our model stems from the sum of a pair of binomial random variables; it is well established that the binomial can be well approximated by the normal under a wide range of parameterizations.[44] It is therefore reasonable to think in terms of an additive normal error term, which is helpful. The bad news is that the variances of the binomial random variables—and the variance of their sum—is not in general uniform. We will discuss the latter problem, and the steps we have taken to deal with it, in a later section of this appendix; for now, we will leave it to one side.

Two consequences of this modeling approach are now apparent. The first is simply that we have found an explanation for the autocorrelation that we have observed in the FMC data. We can generally expect

---

43. The analogous deterministic system has a steady state which is exactly that given by equation 16, so in that sense FMC_{ss} can be thought of as a steady state.

44. For an example, see Mansfield (1983).

an AR(1) process to exhibit first order autocorrelation, and we have detected that in both the F–14 and F/A–18 FMC series.[45]

The other interesting consequence of using this method is that we can produce estimates of both P_1 and P_2. By estimating equation 17, we can get estimates for both FMC_{ss} and α; since each of those is a function of P_1 and P_2, we have two equations in two unknowns and so can solve for P_1 and P_2. The reason we can do this is that we are making use of both the series mean and its autocorrelation—the two together contain more information than a simple mean. We can also think of it as making use of both the series mean and the order in which the observations occur.

Although the connection between Markov processes and first order autocorrelation is fairly well known, the idea of estimating parameters of a Markov chain from an aggregate measure hasn't received much attention. However, we have found one reference—Basawa and Prakasa Rao (1980)—that describes a model that is virtually the same as the one we present here. Thus we have found at least one citation that confirms the validity of our approach.

A final note: there is a correspondence between this type of model and another class of estimable models. Survival or "hazard" models are used widely in engineering and the social sciences in situations where there is a relationship between explanatory variables and duration in a state or time to failure. The underlying structure of such models is frequently similar to, and sometimes the same as, the one we assume for the Markov chain model. We are not using that approach here because, to apply a hazard model empirically, we need data on the individual units, i.e., planes. Think of our Markov chain model as a second-best approach for dealing with occasions such as this one where plane-level data are not available.

45. The order of the autocorrelation has been established in two ways using various SAS routines. First, we examined plots of the relevant partial autocorrelation functions. Then, as a check, we conducted a backward elimination of insignificant lag terms. The evidence from both approaches was unambiguous—the processes are first order.

Incorporating explanatory variables

If this effort is to have any policy relevance, explanatory variables must be included in the model. The form of the model makes this somewhat more complicated than is usually the case in a linear model, since the variables of interest are explaining not the ultimate dependent variable—the FMC rate—but rather are affecting the entry and exit probabilities to and from FMC (P_1 and P_2). Thus we are thinking in terms of treating P_1 and P_2 as functions of the explanatory variables, with the FMC rate in turn being a function of P_1 and P_2.

The central question here is, what functional form should be assumed for the relationships between the explanatory variables—henceforth, the x's—and P_1 and P_2? We can specify a simple linear form as follows:

$$P_i = \beta_i x, \; i = 1, 2 \; . \qquad (18)$$

This formulation is relatively easy to work with. Specifically, it can be shown that the resulting equation for FMC rates can be estimated using ordinary least squares (see Trost 1996). Thus far, we have not attempted to use this linear form, because it allows for the possibility of probabilities outside of the 0 - 1 range. However, it would probably be worthwhile to try it at some point in the future—if only as an alternative way of developing starting values for the nonlinear model which we will now describe.

We have chosen to assume a logistic relationship between P_i and the explanatory variables. This takes the following form:

$$P_i = \frac{e^{\beta_i x}}{1 + e^{\beta_i x}} \; . \qquad (19)$$

Our reason for doing this is the usual one: We wanted a functional form that constrains the predicted P_i's to fall between 0 and 1. The FMC rate equation then becomes

$$FMC_{t+1} = \left(1 - \frac{e^{\beta_1 x}}{1 + e^{\beta_1 x}}\right) FMC_t + \frac{e^{\beta_2 x}}{1 + e^{\beta_2 x}} (N - FMC_t) + \varepsilon_{t+1} \quad . \tag{20}$$

Use of this form has some important implications for estimation methods which we discuss below. Meanwhile, we end this section with two observations concerning the x's. The first is that the notation we have used should be interpreted to include a constant (i.e., β_{0i}). The other is that we have not subscripted the x's, on the principle that any given variable could help explain either or both of the P_i's. Although we have found in practice that no variable seems to have significant explanatory power in both equations, in principle, one could.

Heteroscedasticity

The binomial random variables that appear in equation 14 have variances that are functions of both arguments—$\mathrm{Var}[Bi(x,p)] = xp(1-p)$. Since the observed FMC rate will vary from one period to another, this means that, in general, the variances of the error terms will vary also. We have accommodated this by weighting the observations according to the sum of the calculated variances, where the calculated variances are based on the actual FMC rate and number of aircraft at time t as well as the estimated values of P_1 and P_2. Since the estimation method we use is an iterative numerical approach, this means that the weights are recalculated and reapplied with every iteration.

Varying numbers of aircraft

In the discussion thus far, we have assumed that the number of aircraft is fixed. The reality is that the number of aircraft has varied over time. There are two quite different reasons for this variability:

- Introduction of new and retirement of old aircraft TMS
- Entry into or exit from reporting.

The first of these reasons is normal and needs no elaboration, but the second bears some discussion.

Under certain circumstances, a plane that is no longer fully mission capable need not be counted for the purposes of calculating a squadron's FMC rate. For the purposes of our modeling efforts, the implication of this is that, in reality, a plane can be in a third state: not reporting. We did not have data on this particular state at hand, however, so we had to find some way to cope with movements into and out of reporting. We do have information on the number of aircraft on hand, which makes it possible to deal with this matter in an admittedly oversimplified way.

Our method for coping with changes in the number of aircraft was to make an assumption about what state a plane is in prior to leaving reporting status, and make a similar assumption about the state of a plane that returns to reporting status. In principle, it makes sense to suppose that a plane returning to reporting status is FMC, while one that leaves is perhaps not FMC prior to leaving. This is because the squadron's leaders wish to report as high an FMC rate as possible; they will therefore want to have any available FMC plane in reporting status. Unfortunately, our ability to specify these sorts of assumptions was limited because of limitations in the software we were using. This forced us to treat entering and exiting aircraft in the same way. We experimented with assuming that entering and exiting aircraft were all FMC, all not FMC, or FMC in proportion to the state of the rest of the squadron. We got the best results with the assumption that all aircraft leaving and entering reporting were FMC the month before—this became our working assumption. The resulting equation we estimated was

$$FMC_{t+1} = \left(1 - \frac{e^{\beta_1 x}}{1 + e^{\beta_1 x}}\right) FMC_t + \frac{e^{\beta_2 x}}{1 + e^{\beta_2 x}}(N - FMC_t) \quad + \\ \Delta Number(a/c) + \varepsilon_{t+1} \ . \tag{21}$$

This approach is arguably too simplistic; it may be possible in the future to employ a more sophisticated approach such as working with a truly multinomial—rather than binomial—structure.

Estimation details

We estimated equation 21 using SAS *Proc Nlin* with the Marquardt search option, step halving switched off, and iterative reweighting to deal with the heteroscedasticity described earlier. We initialized parameter estimates at zero, though we did experiment with other starting points. This often resulted in the procedure failing to converge. In general, nonconvergence was a problem.

Proc Nlin reports asymptotic 95-percent confidence intervals for the parameter estimates. We used these as the basis for our conclusions concerning statistical significance. However, these results should be treated with caution. The iterative reweighting of the observations means that what is being minimized is not—strictly speaking—least squares, so the standard errors may not be valid even asypmtotically. We have nonetheless taken these as a benchmark—mainly because we have no workable alternative. Further investigation may validate the use of these intervals, or we may find another way around the problem. Basawa and Prakasa Rao (1980) propose a different way of handling the heterosedasticity that may resolve this issue. In the meantime, the reader should keep in mind that the significance tests are suspect.

The variables that we found to be significant are listed in the main body of this report. Here we list other variables we tried whose parameter estimates we found to be not significant:

- Utilization rates (highly correlated with sorties/plane)
- Personnel SORTS (correlated with the PQI)
- Officer quality (as measured by flight school scores)
- Enlisted turnover
- Carrier supply SORTS (for deployed squadrons)
- Response time for spares requests (highly correlated with the percentage of repairs accomplished in one day)
- Months since last deployment.

A note about months since last deployment—we used both it and its square to capture the notion that nondeployed squadrons would experience a trough in readiness and then rebound as the next deployment approaches. Although the parameter estimates were small and insignificant, they were of the signs that we expected and indicated that the bottom of the trough would occur at about 10 months from the last deployment, which seems about right.

Conclusion

The Markov model has some disadvantages—it is unwieldy to work with and raises some subtle statistical issues. However, it does appear to capture the character of the data and the underlying process by which it is generated. It can also explain the autocorrelation in the data in a way that a conventional linear model cannot. We consider it to be a promising but as yet unproved alternative method for analyzing and forecasting readiness data.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C: The personnel quality index (PQI)

The enlisted personnel quality index is calculated in the exact same manner as it was in Junor, 1996. We refer readers to that research for specific technical details.

The means by which the index is calculated is, again, principal component analysis. We applied the technique to the following variables:

- The percentage of the squadron with a high school degree
- The percentage testing in the upper mental group (categories I, II and IIIA) on the Armed Forces Qualification Test (AFQT)
- The percentage of the squadron demoted that month
- The percentage of the squadron who were promoted to E5 or better with less than 4 years of experience
- The average length of service of the squadron's enlisted personnel.

We computed an index two ways. First, for each squadron, by month, from 1982 through 1995. This index is used as a variable in our readiness equations. We also computed an index for average squadrons, monthly, from 1982 through 1995. This is the version shown in the diagram in the text. The results are very similar.

The time series index captures about 85 percent of the information available in the data, whereas the panel version explains about 51 percent.[46] Table 16 lists the weights.

---

46. It is always more difficult to explain panel data than averaged, time series data. Think of it as the difference between having to explain what a specific steadier is doing in a given month versus what the average squadron is doing for the same month. There's a lot more random noise happening on the panel in individual level.

Table 16. Weights for the personnel quality index

Variable	Time series weight	Panel weight
The percentage with a high school degree	.45	.51
The percentage in the upper mental group	.43	.38
The percentage demoted	-.42	-.28
The percentage of rapid advancements	-.47	-.51
The average length of service	.46	.50
Explanation of available information	85%	51%

Appendix D: A closer look at the data—variable definitions and summary statistics

The data we used for this analysis are for individual VA, VF, and VFA squadrons and go back monthly from 1982 through 1995. The data sources include the Enlisted Master Record (EMR), the Officer Master Record (OMR), the Ship Information Digest (SID), the Aircraft Information Database (AID), the Naval Sea Logistics Center, the SORTS database, and the ship employment history database. Here we describe each variable we use in the modeling and provide summary statistics for them.

Variable definitions

We use the first set of variables as dependent variables in our equations. They each provide a measure of readiness in one of the resource areas.

- Personnel: This variable is derived using SORTS data. It is calculated as the fraction of the month that a squadrons is C1 for personnel.
- Training: This is also a SORTS variable. Like personnel, it is calculated as the fraction of the month that a squadron is C1 for training.
- FMC: This variable refers to the percentage of time the squadron's aircraft are able to complete all their missions.
- Flight hours between failures: Here we count the number of flight hours flown per squadron per month and divide by the total number of reported failures.
- AIMD repairs: For this variable, we calculated the probability that all items processed through the AIMD were actually repaired at the AIMD.

This set of variables measures aspects of a squadron's deployment cycle:

- Deployed indicator: Value is equal to 1 if the squadron is deployed that month, 0 otherwise.
- Deployment history: The number of months, out of the last 6, that the squadron was deployed.
- Months since last deployed: Value is equal to the number of months since the squadron was last deployed. Value is equal to 0 if the squadron is currently deployed.

These variables measure the rate at which aircraft are flown:

- Flight hours: The average number of flight hours flown by a squadron in a given month.
- Utilization rate: The number of flight hours divided by the number of aircraft in that squadron. This is calculated monthly.
- Sorties: The number of sorties flown by a squadron per month divided by the number of aircraft.

The following list describes manpower variables:

- Personnel quality: This is the value of the personnel quality index described in appendix C. The measure is for enlisted personnel per squadron per month.
- Enlisted manning: This is a weighted sum of enlisted personnel in each pay grade, weighted by pay. The sum is divided by the number of billets for that squadron.
- Number of officers: This is the number of officers assigned to a squadron divided by the number of reporting aircraft.
- Officer turnover (3 month): This variable measures the number of officers present for any given squadron that were not there 3 months ago.
- Officer turnover (6 month): The same as above, only looking back 6 months.

- Enlisted turnover (3 month): This variable is calculated exactly the same as the officer 3-month turnover, but this version counts enlisted personnel.

These are our set of supply-related variables:

- Consumables on hand: The percentage of time that requests for consumables are filled within a day.
- Repairables on hand: The percentage of time that requests for repairables are filled within a day.
- Response time for consumables: The average number of days required to fulfill a consumables request.
- Response time for repairables: The average number of days required to fulfill a repairables request.

Summary statistics

The tables below list descriptive statistics for all the variables used in our modeling. For each, we have provided a mean, standard deviation, minimum, and maximum. Table 17 lists the statistics for our dependent variables; whereas table 18 lists the statistics for our independent variables. Finally, table 19 indicates the mix of aircraft types in our data. All variables are computed per squadron per month unless otherwise specified.

Table 17. Summary statistics for dependent variables

Statistic	Personnel	Training	Flight hours between failures	AIMD repairs	FMC
Mean	45.3	42.3	.86	56.7	60.2
Minimum	13.0	0.0	0.05	50.1	36.3
Maximum	85.4	73.6	1.28	60.8	75.1
Mean occurs:	Jul-89	Oct-83	Jun-90	May-93	Oct-91
Minimum occurs:	Jan-82	May-82	Jan-83	Dec-95	Sep-82
Maximum occurs:	Oct-95	Nov-87	Mar-93	Dec-89	Apr-95

Table 18. Summary statistics for model independent variables

Variable	Mean	Standard deviation	Minimum	Maximum
Deployed indicator	.32	.47	0	1
Months since last deployment	6.5	7.3	0	46
Number of months (out of past 6) squadron has been deployed[a]	1.9	2.2	0	6
Number of aircraft	10.1	2.3	0.1	24.5
Average age of aircraft TMS	14.5	6.9	0	28
Sorties per aircraft	19.4	8.4	0	161
Deployed sorties per aircraft	6.3	10.2	0	161
Flight hours	331.2	153.7	1	1463
Flight hours per aircraft (utilization)	33.4	14.4	0.10	143
6-month accrual of FMC flight hours[b]	118,287	46,753	616	328,830
Number of officers, 6-month average[a]	2.6	1.4	0.82	25.9
Enlisted manning relative to billets	93	8.7	8.3	141
Officer 3-month turnover	9.4	7.7	0	78.6
Officer 6-month turnover	18.6	10.5	0	81.3
Enlisted 3-month turnover	9.7	4.5	0	67.3
Enlisted personnel quality (squadron)	-0.003	1.00	-3.1	2.9
Enlisted personnel quality (ship)	-0.002	0.59	-4.3	1.2
Probability of having consumables on hand	73.1	22.8	0	100
Probability of having repairables on hand	76.0	22.5	0	100
Average response time for consumables	8.9	22.7	0	309.7
Average response time for repairables	7.4	21.3	0	365.0
Number of days per month ship is C1 for equipment	5.3	11.0	0	31
Total number of days ship is C1 for equipment over last 6 months	31.4	49.6	0	185

a. These variables represent computations spanning 6 months. These variables were used in our training equation.

b. To calculate this variable, we first multiply the FMC percentage (ranges from 0 to 100) by the number of flight hours per squadron per month. Then we summed these products over the last 6 months. Therefore, the average number of FMC flight hours for a squadron for a month would be 197.

Table 19. Aircraft mix

Type	Frequency	Percent
A–6E	2,022	22.1
A–7E	1,708	18.6
F–4S	102	32.6
F–14A	2,984	32.6
F–14B	195	2.1
F–14D	89	9.7
F/A–18A	885	9.7
F/A–18C	1,177	12.8

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

Ishwar V. Basawa and B. L. S. Prakasa Rao. *Statistical Inference for Stochastic Processes.* London: Academic Press, 1980

Linda C. Cavalluzzo and Stanley A. Horowitz. *Resources-to-Ship Training Readiness Study*, Jul 1987. Resource Consultants, Inc.

William H. Greene. *Econometric Analysis.* New York, NY: Macmillan Publishing Company, 1993

Laura J. Junor and Jessica S. Oi. *A New Approach to Modeling Ship Readiness*, Apr 1996 (Center for Naval Analyses Research Memorandum 95-239)

Edwin Mansfield. *Statistics for Business and Economics—Methods and Applications.* New York, NY: W. W. Norton, 1983

Matthew T. Robinson, et al. *Avoiding a Hollow Force: An Examination of Navy Readiness*, Apr 1996 (Center for Naval Analyses Research Memorandum 95-238)

Robert Trost. *Estimating Hazard Rates When Duration Data Are Not Available*, unpublished memorandum, 12 Dec 1996

THIS PAGE INTENTIONALLY LEFT BLANK

List of figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of tables

Distribution list

Research Memorandum 97-14

SNDL

21A1	CINCLANTFLT NORFOLK VA
21A2	CINCPACFLT PEARL HARBOR HI
21A3	CINCUSNAVEUR LONDON UK
22A1	COMSECONDFLT
22A2	COMSEVENTHFLT
22A2	COMTHIRDFLT
22A3	COMSIXTHFLT
24A1	COMNAVAIRLANT NORFOLK VA
24A2	COMNAVAIRPAC SAN DIEGO CA
24D1	COMNAVSURFLANT NORFOLK
24D2	COMNAVSURFPAC SAN DIEGO
24H1	COMTRALANT NORFOLK VA
A1A	SECNAV WASHINGTON DC
A1B	UNSECNAV
A1H	ASSTSECNAV MRA WASHINGTON
A2A	OPA
A2A	USACOM
A5	BUPERS
A5	PERS-2
A6	PERS-4
A5	PERS-5
A6	HQMC ACMC
A6	HQMC AVN
A6	HQMC CMC
A6	HQMC I&L
A6	HQMC MPR & RA
A6	HQMC P&R
A6	HQMC PP&O
A6	ASN(I&E)
B1A	DEPSECDEF
	Attn: LOUIS FINCH (2 COPIES)
B1B	SECDEF
	Attn: JOE ANGELLO (2 COPIES)
B1B	PA&E
V12	CG MCCDC - ATTN: CMCC
	Attn: COMMANDING GERERAL
	Attn: STUDIES AND ANALYSES DIVISION

OPNAV

N00
N00K
N09
N122C2D
N3/N5
N31/N52
N4
N41
N414
N42
N43
N4B
N8
N80
N81
  Attn: 2 COPIES
N813
N814
  Attn: 2 COPIES
N85
N85B
N86
N88

OTHER

IDA
LMI
RAND SANTA MONICA
RAND WASHINGTON

CNA